

HUMBOLDT-UNIVERSITÄT ZU BERLIN
SCHOOL OF BUSINESS AND ECONOMICS
LADILAUS VON BORTKIEWICZ CHAIR OF STATISTICS



Analyse der Wahlergebnisse der Bundestagswahlen anhand der Strukturdaten der Wahlergebnisse

Analysis of Results of the Bundestag Elections using Structure Data from the Election Districts

Bachelorarbeit
VORGELEGT VON
Viet Huong Pham
(538820)

ZUR ERLANGUNG DES AKADEMISCHEN GRADES
BACHELOR OF SCIENCE
IN VOLKSWIRTSCHAFTSLEHRE



PRÜFER
Prof. Dr. Wolfgang Härdle
Dr. Sigbert Klinke
BERLIN, 26. JANUAR 2015

Inhaltsverzeichnis

1	Einleitung	1
2	Methodik	3
2.1	Theorie.....	3
2.2	Datensatz.....	6
2.3	Spezifikation.....	8
2.4	Regression und Prüfung.....	13
3	Regression.....	20
3.1	Prüfung der Voraussetzungen	20
3.2	Regression	25
3.3	Modellprüfung	28
3.4	Prüfung der Regressionskoeffizienten	31
4	Schluss.....	35
5	Literatur	37
6	Anhang.....	40

Abkürzungsverzeichnis

<i>Age1</i>	Altergruppe1: Anteil der unter 18-Jährigen
<i>Age2</i>	Altergruppe2: Anteil der 18 bis 25-Jährigen
<i>Age3</i>	Altergruppe3: Anteil der 25 bis 35-Jährigen
<i>Age4</i>	Altergruppe4: Anteil der 35 bis 60-Jährigen
<i>Age5</i>	Altergruppe4: Anteil der unter 18-Jährigen
<i>bpb</i>	Bundeszentrale für Politische Bildung
<i>BTW</i>	Bundestagswahl
<i>CDU/SU1</i>	Erststimmenanteil der CDU/CSU
<i>CDU/SU2</i>	Zweitstimmenanteil der CDU/CSU
<i>Edu1</i>	Bildungsgruppe1: Anteil der Schulabgänger ohne (Haupt-) Schulabschluss
<i>Edu2</i>	Bildungsgruppe2: Anteil der Schulabgänger mit Hauptschulabschluss
<i>Edu3</i>	Bildungsgruppe3: Anteil der Schulabgänger mit Realschulabschluss
<i>Edu1</i>	Bildungsgruppe4: Anteil der Schulabgänger mit allgemeiner Hochschulreife
<i>FDP1</i>	Erststimmenanteil der FDP
<i>FDP2</i>	Zweitstimmenanteil der FDP
<i>GRÜ1</i>	Erststimmenanteil der Grünen
<i>GRÜ2</i>	Zweitstimmenanteil der Grünen
<i>LIN1</i>	Erststimmenanteil der Linken
<i>LIN2</i>	Zweitstimmenanteil der Linken
<i>Mig</i>	Anteil der nicht deutschen Bevölkerung
<i>Sec1</i>	Anteil der sozialversicherungspflichtigen in Land-, Forstwirtschaft und Fischerei
<i>Sec2</i>	Anteil der sozialversicherungspflichtigen im produzierenden Gewerbe
<i>Sec3</i>	Anteil der sozialversicherungspflichtigen in Handel, Verkehr und Nachrichtendienst
<i>Sec4</i>	Anteil der sozialversicherungspflichtigen im Dienstleistungsbereich
<i>Sex</i>	Anteil der männlichen Bevölkerung
<i>SPD1</i>	Erststimmenanteil der
<i>SPD2</i>	Zweitstimmenanteil der
<i>Une</i>	Arbeitslosenquote

Abbildungsverzeichnis

ABBILDUNG 3.1: STREUDIAGRAMM ZUR UNTERSUCHUNG DER LINEAREN BEZIEHUNG	22
ABBILDUNG 3.2: HISTOGRAMM ZUR UNTERSUCHUNG DER RESIDUEN AUF NORMALVERTEILUNG BEI REGRESSION DER ZWEITSTIMMEN CDU/CSU 2013.	23
ABBILDUNG 3.3: P-P-DIAGRAMM ZUR UNTERSUCHUNG DER RESIDUEN AUF NORMALVERTEILUNG BEI REGRESSION DER ZWEITSTIMMEN CDU/CSU 2013.	23
ABBILDUNG 3.4: HISTOGRAMM ZUR UNTERSUCHUNG DER RESIDUEN AUF NORMALVERTEILUNG BEI REGRESSION DER ERSTSTIMMEN DER GRÜNEN 2013.	23
ABBILDUNG 3.5: HISTOGRAMM ZUR UNTERSUCHUNG DER RESIDUEN AUF NORMALVERTEILUNG BEI REGRESSION DER ERSTSTIMMEN DER GRÜNEN 2013.	24
ABBILDUNG 3.6: STREUDIAGRAMM ZUR PRÜFUNG DER RESIDUEN AUF HOMOSKEDASTIZITÄT BEI REGRESSION DER ZWEITSTIMMEM DER SPD.	25

Tabellenverzeichnis

TABELLE 3.1: TEST AUF MULTIKOLLINEARITÄT ZWISCHEN BESTIMMTEN GRUPPEN UNABHÄNGIGER VARIABLEN DURCH LINEARE REGRESSION	21
TABELLE 3.2: GÜTE DER AUERWÄHLTEN MODELLE FÜR DIE BUNDESTAGSWAHLEN 2002, 2005, 2009 UND 2013.	30
TABELLE 7.1: TEST AUF MULTIKOLLINEARITÄT MIT SPEARMAN-KORRELATIONSKOEFFIZIENTEN.....	40
TABELLE 7.2: TEST AUF MULTIKOLLINEARITÄT MIT SPEARMAN-KORRELATIONSKOEFFIZIENTEN.....	41
TABELLE 7.3: AUERWÄHLTE MODELLE FÜR DIE 15. BUNDESTAGSWAHL (2002) DER WAHLBETEILIGUNG, ERSTSTIMME CDU/CSU UND ZWEISTIMME CDU/CSU.	42
TABELLE 7.4: AUERWÄHLTE MODELLE FÜR DIE 15. BUNDESTAGSWAHL (2002) DER ERST- UND ZWEISTIMME SPD.	43
TABELLE 7.5: AUERWÄHLTE MODELLE FÜR DIE 15. BUNDESTAGSWAHL (2002) DER ERST-UND ZWEISTIMME FDP UND ERSTSTIMME DER LINKEN.	44
TABELLE 7.6: AUERWÄHLTE MODELLE FÜR DIE 15. BUNDESTAGSWAHL (2002) DER ZWEISTIMME DER LINKEN, ERST- UND ZWEISTIMME DER GRÜNEN.	45
TABELLE 7.7: AUERWÄHLTE MODELLE FÜR DIE 16. BUNDESTAGSWAHL (2005) DER WAHLBETEILIGUNG UND ERSTSTIMME DER CDU/CSU	46
TABELLE 7.8: AUERWÄHLTE MODELLE FÜR DIE 16. BUNDESTAGSWAHL (2005) DER ZWEISTIMME DER CDU/CSU UND ERSTSTIMME DER SPD.	47
TABELLE 7.9: AUERWÄHLTE MODELLE FÜR DIE 16. BUNDESTAGSWAHL (2005) DER ZWEISTIMME DER SPD UND ERST- UND ZWEISTIMME DER FDP.	48
TABELLE 7.10: AUERWÄHLTE MODELLE FÜR DIE 16. BUNDESTAGSWAHL (2005) DER ERST- UND ZWEISTIMME DER LINKEN.	49
TABELLE 7.11: AUERWÄHLTE MODELLE FÜR DIE 16. BUNDESTAGSWAHL (2005) DER ERST- UND ZWEISTIMME DER GRÜNEN.....	50
TABELLE 7.12: AUERWÄHLTE MODELLE FÜR DIE 17. BUNDESTAGSWAHL (2009) DER WAHLBETEILIGUNG, ERST- UND ZWEISTIMME CDU/CSU.	51

TABELLE 7.13: AUSERWÄHLTE MODELLE FÜR DIE 17. BUNDESTAGSWAHL (2009) DER ERST- UND ZWEITSTIMME SPD UND ERSTSTIMME FDP.	52
TABELLE 7.14: AUSERWÄHLTE MODELLE FÜR DIE 17. BUNDESTAGSWAHL (2009) DER ZWEITSTIMME FDP UND ERSTSTIMME DER LINKEN.	53
TABELLE 7.15: AUSERWÄHLTE MODELLE FÜR DIE 17. BUNDESTAGSWAHL (2009) DER ZWEITSTIMME DER LINKEN, ERST UND ZWEITSTIMMEN DER GRÜNEN.	54
TABELLE 7.16: AUSERWÄHLTE MODELLE FÜR DIE 18. BUNDESTAGSWAHL (2013) DER WAHLBETEILIGUNG, ERST- UND ZWEITSTIMMEN DER CDU/CSU.	55
TABELLE 7.17: AUSERWÄHLTE MODELLE FÜR DIE 18. BUNDESTAGSWAHL (2013) DER ERST- UND ZWEITSTIMMEN DER SPD.	56
TABELLE 7.18: AUSERWÄHLTE MODELLE FÜR DIE 18. BUNDESTAGSWAHL (2013) DER ERST- UND ZWEITSTIMMEN DER FDP UND ERSTSTIMME DER LINKEN.	57
TABELLE 7.19: AUSERWÄHLTE MODELLE FÜR DIE 18. BUNDESTAGSWAHL (2013) DER ZWEITSTIMME DER LINKEN ERST- UND ZWEITSTIMMEN DER GRÜNEN.	58

1 Einleitung

Die vorliegende Arbeit untersucht anhand statistischer Methoden die Beziehung zwischen sozio-ökonomischen Rahmenbedingungen in Wahlkreisen und den Ergebnissen der Bundestagswahlen von 2002 bis 2013. Hierbei werden die vom Bundeswahlleiter erhobenen Strukturdaten und Bundestagswahlergebnisse der Wahlkreise zur Erstellung eines statistischen Modells angewendet, um ein rein mathematisches Ergebnis zu formulieren.

Die Auseinandersetzung mit diesem Thema kann sowohl für Politik- und Sozialwissenschaftler als auch für Volkswirte von Interesse sein. Es liegen derzeit wenige Untersuchungen bzgl. der Einflüsse auf Wahlergebnisse vor. Im Hinblick auf die Bundestagswahlen 2013 stellte sich beispielsweise Dr. Katrin Weller (2014) die Frage, ob Twitter und Facebook das Wahlergebnis vorhersagen können.¹ Ähnlich Martin Beckmann, der sich 2014 mit dem Thema „Lässt sich der Ausgang von politischen Wahlen durch die Analyse von Twitterdaten vorhersagen?“, befasst hat. Im Jahr 2009 beschäftigte sich Prof. Dr. Ulrich Rosar auf empirischer Art und Weise mit der „Bedeutung der physischen Attraktivität von Spitzenkandidaten für den Wahlerfolg ihrer Parteien“. Zuvor wurden in anderen Studien ähnliche Problematiken in Bezug auf Wahlkreiskandidaten bzw. Listenkandidaten behandelt². Eine ausführliche Darlegung von Alf Minzel und Stefan Immerfalls (1994) „Zusammenhangsanalyse von lokaler Mitgliederpräsenz, Wahlergebnis und Sozialstruktur“, liefert Susanne Koch in ihrem Buch „Parteien in der Region“. Eine besondere Verbindung der genannten Literatur zur vorliegenden Arbeit ist nicht nur die Thematik sondern auch das dazu verwendete Mittel: Die multivariate oder auch multiple Regressionsanalyse. Lediglich die räumliche Eingrenzung und die dazu einbezogenen Strukturdaten weichen ab. Im Unterschied zum Buch, wird diese Studie nicht ausschließlich Wahlergebnisse Bayerns betrachten, sondern die der Bundesrepublik und hat daher eine allgemeingültige Aussagekraft. Die Bundeszentrale für politische Bildung (bpb) untersuchte 2010 die Wählerlandschaft der einzelnen Parteien bezüglich Geschlecht und Altersstruktur.

Am Anfang dieser Arbeit wird der theoretische Hintergrund diskutiert. Zum einen wird allgemeine Auskunft über Parteienziele der wahlstärksten Parteien: SPD, CDU und CSU (auch abgekürzt CDU/CSU), FDP, Die Linke (LIN), Bündnis 90/Die Grünen (GRÜ) gegeben. Zum

¹ Twitter and society

² Vgl. Efran/Patterson, 1979; Klein/Rosar 2005; Berggren et. al., 2007; King/Leigh, 2007; Rosar et. al. 2008

anderen werden Strukturdaten analysiert. Der darauf folgende Hauptbestandteil dieser Studie besteht aus einer quantitativ empirischen Ausführung mehrerer Regressionsanalysen mittels SPSS. Die Frage nach der Existenz eines Zusammenhangs zwischen wahlkreisspezifischen Strukturdaten und den Bundestagswahlergebnissen wird detailliert erforscht und beantwortet. Die Wahl der statistischen Modellierung, Durchführung von verschiedenen Tests werden dabei im Vordergrund stehen.

Ziel dieser Analyse ist es, den Zusammenhang zwischen Strukturdaten und Wahlergebnissen der Bundestagswahlen auf quantitative Weise zu beschreiben und zu erklären. Dabei werden Werte der abhängigen Variable Y geschätzt. Mittels der Untersuchung kann die Kausalbeziehung vom prozentualen Stimmenanteil der Parteien $Y = f(X)$ und Strukturdaten X dargestellt werden (Backhaus et al., 2011).

2 Methodik

In diesem Kapitel werden zuerst die vom Bundeswahlleiter und Destatis gegebenen Daten genannt und die dahinter stehende Theorie beschrieben. Nun werden das anzuwendende multiple Regressionsmodell, seine Mechanismen und Annahmen erklärt. Anschließend folgt eine theoretische Darstellung der Methodik zur Prüfung dieser Annahmen zum Modell im Punkt 2.3. Darauf folgend werden im Punkt 2.4 die zur Prüfung des Modells verwendeten Techniken beschrieben. Der Methodikabschnitt wird mit einer Erläuterung der Durchführung einer Prognose abgeschlossen.

2.1 Theorie

Nach Artikel 38 (1) und 39 (1) des Grundgesetzes finden alle vier Jahre die Bundestagswahlen in allgemeiner, unmittelbarer, freier, gleicher und geheimer Art und Weise statt. Bei der Bundestagswahl, dessen Ergebnis in dieser Arbeit untersucht wird, kann jeder Wähler zwei Stimmen abgeben. Dabei werden insgesamt mindestens 598 Abgeordnete in den Bundestag gewählt. Zuzüglich können Überhang- und Ausgleichsmandate entstehen, die sich bei der Verteilung der Sitze ergeben. Die Mandate werden zur Hälfte direkt über die 299 Wahlkreise vergeben, die andere Hälfte wird über die Landeslisten der Parteien bestimmt. Mit der Erststimme wählt der Wähler direkt seinen bevorzugten Regionalvertreter in den Bundestag. Wer die meisten Stimmen bekommt, gewinnt den Sitz. Die Zweitstimme – auch Kanzlerstimme genannt - erlaubt die Wahl einer Partei³. Sie wird als wichtigere Stimme aufgefasst, da sie über das Kräfteverhältnis der Parteien im Bundestag und letzten Endes über Koalitionen und Regierungschef entscheidet. Dies sollte während der empirischen Analyse im Hinterkopf behalten werden. Zum Verständnis der zu untersuchenden These, wird im Folgenden ein kleiner Überblick die Wählerlandschaft der stärksten Parteien Deutschlands: CDU/CSU, SPD, FDP, Die Linke, und Die Grünen gegeben. Als Quelle dienen Veröffentlichungen der Bundeszentrale für politische Bildung (bpb), da sie bereits Untersuchungen zum Thema Wählerschaft durchgeführt hat. Es macht dabei Sinn, Daten seit 2002 zu sammeln und zusammenzufassen, da sich seitdem die Aufteilung und Anzahl der Wahlkreise nicht geändert haben. Der Aspekt der Einhaltung der Wahlprogramme durch die

³ vgl. www.Bundestag.de

Regierung wird außer Acht gelassen. Dies ist durchaus ein Defizit der Studie. Jedoch soll die Annahme verdeutlicht werden, dass jede Partei mittels ihrer Wahlprogramme bestimmte „Zielgruppen“ der Bevölkerung ansprechen. Somit stehen bestimmte Merkmale bzgl. der Bevölkerungsstruktur in Beziehung mit dem Wahlergebnis im jeweiligen Wahlkreis.

Die Feststellung der erklärenden Variablen X und der erklärten Variablen Y erfolgt auf theoretischer Ebene: Das Wahlergebnis wird höchstwahrscheinlich von den sozialen Strukturen in einem Wahlkreis beeinflusst. Generell wird vermutet, dass bestimmte sozio-ökonomische Merkmale in einem Wahlkreis eine Beziehung zum Stimmenanteil einer Partei in diesem Wahlkreis haben. Es wird also angenommen, dass der Ausgang der Bundestagswahl von sozio-ökonomischen Eigenschaften eines Wahlkreises linear abhängig ist und nicht andersherum. Unter diesem Ansatz gelten die Ergebnisse der Bundestagswahlen als abhängige Variable Y . Strukturdaten lassen sich als unabhängige Variable X interpretieren. Ziel der Regressionsanalyse ist es, eine Aussage über die Beziehungen zwischen den beschriebenen Einheiten zu erfassen. In der Regel streuen die Werte der abhängigen Variablen bei konstanten Werten der unabhängigen Variablen. Das folgt aus dem Grund, dass unkontrollierbare Einflüsse auf die abhängige Variable existieren, die nicht aus den spezifizierten X -Variablen folgen. Somit ist das vorliegende Regressionsmodell stochastisch formuliert (vgl. Urban, 1982). Damit kann die Anwendung eines einfachen linearen Regressionsmodells ausgeschlossen werden. Aus diesem Grund eignet sich für die Analyse die Methodik der multiplen bzw. multivariaten Regression. Die Theorie dazu folgt später. Zunächst wird ein kurzer Überblick über die erklärten und erklärenden Variablen vom Jahr 2002 bis 2013 gegeben.

Bei einer Wahluntersuchung ist die Analyse der Wahlbeteiligung besonders interessant. Dem möglichen Bestehen einer Beziehung zwischen den Strukturdaten und Wahlergebnissen wird hier natürlich auch nachgegangen werden. Die unabhängigen Variablen sind Anteile der Wahlberechtigten, die sich aus dem Quotienten der absoluten Wähler und der Wahlberechtigten berechnet wird. Des Weiteren befasst sich diese Studie auch mit dem Ergebnis der einzelnen Parteien. Die Kalkulation der Anteile der Wählerschaft einer Partei geschieht durch die Division der absoluten Anzahl der erhaltenen gültigen Erst- oder Zweitstimmen durch die Anzahl der gültigen Stimmen im gesamten Wahlkreis. Nun wird anhand Untersuchungen der bpb genauer diskutiert, welche Partei welche Wählerschaft anspricht.

Seit der Ost-West-Vereinigung der Christlich Demokratischen Union (**CDU/CSU**) 1990 änderten sich die Ziele der Partei kaum (vgl. Olzog, 1999). Noch heute steht der Grundsatz der „Freiheit und Verantwortung“ im Vordergrund und wurde seit 2007 in „Freiheit, Solidarität und Gerechtigkeit“ umformuliert. Die christlich-sozial, liberal und wertkonservativ geprägte Organisation nennt sich „Volkspartei der Mitte“ und versucht sich an alle Bevölkerungsschichten zu wenden. Der Bundeszentrale für politische Bildung (2010) zu Folge, ist die Wählerschaft der CDU heterogen. Den Untersuchungen nach, stimmen immer mehr jüngere Arbeiter für die CDU (vgl. Schneiker, 2010). Bei der Untersuchung wird für die Wahlkreise Bayerns statt Stimmen der CDU die der CSU einberechnet.

In den letzten zwei Wahljahren wurde eine Senkung des Anteils der arbeitenden Wählern der Sozialdemokratischen Partei Deutschlands (**SPD**) beobachtet. Obwohl ein großer Teil dieser Gruppe Stimmengabe für die SPD blieb, ist laut bpb ein Anstieg bei den Angestellten und Beamten unter den Wählern zu erkennen. Der Stimmanteil aus dem aufstiegsorientierten und dem intellektuellen Milieu nimmt indessen zu. Beamte, Akademiker, Angestellte mit mittleren und höheren Schulabschlüssen und Selbstständige gehören der genannten Gruppe an. Der Wähleranteil mit „eher unterdurchschnittlichen Bildungsstand und einfacherer Beschäftigung“, die eins die Stammwählergruppe der Partei bildeten, schwindet hingegen. Zudem konnten große Wahlerfolge der SPD im Norden der Republik nachgewiesen werden (vgl. Mertens, 2010).

Die Wählerschaft der Freien Demokratischen Partei (**FDP**) verlagerte sich während der sozialliberalen Koalition vom protestantischen Mittelstand auf die Mittelschicht, zu der sich Angestellte und Beamte zuordnen lassen. Derzeit findet die FDP ihre Stammwähler bei Führungskräften und Selbstständigen in der Industrie und Wirtschaft und bei Bürgern mit hohem Bildungsgrad (vgl. Kortmann, 2010).

Nach der bpb hat **Die Linke (LIN)** ihre Wählerschaft aus Zeiten der Linkspartei PDS mit Schwerpunkt in den neuen Bundesländern beibehalten. „Im Vergleich zu den anderen Parteien hatten die Wähler der Vorgängerpartei PDS in den 1990er-Jahren ein formal höheres Bildungsniveau: 26 Prozent der damaligen PDS-Wählerschaft hatten ein abgeschlossenes Hochschulstudium, im Vergleich zu elf Prozent bei den übrigen Parteien“ (Vgl. Taken, 2010).

„Jung, weiblich, gut gebildet“ – so beschreibt Autor der bpb Serkan Agci (2010) das allgemeine Bild der Wähler von **Die Grünen (GRÜ)**. Mit der andauernden Thematik der Nachhaltigkeit in Wirtschaft und Energieerzeugung gewinnt sie die Stimmen der Altersgruppe der 18- bis 25-Jährigen. Diese sind über die Zeit „gemeinsam mit der Partei älter geworden“ (ebenda, 2010). Mit der Erweiterung ihrer politischen Ziele, die in den letzten Jahren auch außerhalb des Umweltaspekts liegen, gewinnt die Partei zunehmende Beliebtheit bei der älteren Wählergruppe.

2.2 Datensatz

Nun werden auch die unabhängigen Variablen dargelegt. Der Bundeswahlleiter veröffentlicht zu jeder Bundestagswahl eine Reihe an Strukturdaten über die einzelnen Wahlkreise. Die sich eignenden Prädikatoren für das Modell lassen sich in einer Reihe von Untersuchungen erkennen, die später erfolgen. Dabei werden nur diejenigen Prädikatoren gewählt, die für jedes Wahljahr erhoben wurden und hinter denen sich eine Beziehung zu Y vermuten lässt. Eine Ausnahme bildet das Wahljahr 2009, wo einige Angaben über die sozialversicherungspflichtigen Beschäftigten in den verschiedenen Branchen nicht gegeben sind. Aus diesem Grund können keine Beobachtungen bezüglich den Anteilen der versicherungspflichtigen Beschäftigten in den Sektoren in das Modell einfließen. Alle anderen potentiellen Regressoren können jedoch für das Wahljahr 2009 eingebracht werden. Sämtliche Prädikatoren, die theoretisch in Beziehung zu den Wahlergebnissen stehen, werden vorgestellt. Eine Beurteilung dieser, wo anschließend eine Auswahl der relevanten Daten stattfindet, folgt im Punkt 3.1.

Prädiktor Geschlecht (*sex*), der die Anteile der männlichen Bevölkerung angibt, ist nicht direkt gegeben. Dieser lässt sich jedoch leicht berechnen, da die Anzahl der Gesamtbevölkerung und die der männlichen gegeben sind.

Regressor Migration (*mig*), besteht aus dem Anteil der nicht-deutschen Bevölkerung in den Wahlkreisen. Diese ist nicht direkt vom Bundeswahlleiter gegeben, sondern muss aus den gegebenen Daten: Anzahl der Bevölkerung und Anzahl der deutschen Bevölkerung aus dem jeweiligen Wahlkreis berechnet werden. Für das Wahljahr 2002 werden Zahlen von 1999 und für die anderen Wahljahre jeweils unmittelbar aus dem vorherigen Jahr einbezogen.

Der nächste Regressor Arbeitslosigkeit (*une*) thematisiert mittels Arbeitslosenquote vom Vorjahr eine Kausalität zwischen Arbeitslosigkeit und den Stimmenanteilen der Bevölkerung für eine Partei. Für die Arbeitslosenquote wird in Bezug auf die abhängigen zivilen Erwerbspersonen berechnet:

$$\text{Arbeitslosenquote} = \frac{\text{Arbeitslose}}{\text{abhängige zivile Erwerbstätige} + \text{Arbeitslose}} * 100.^4 \quad (1)$$

Zu arbeitslosen Personen werden Arbeitsuchende bis zu einem Alter von 65 Jahren gezählt, welche beschäftigungslos sind oder pro Woche weniger als 15 Stunden arbeiten und Beschäftigung suchen. „Abhängige zivile Erwerbspersonen“ sind sozialversicherungspflichtig und geringfügig Beschäftigte, Personen in Arbeitsgelegenheiten, Beamte und Arbeitslose.

Die Regressoren Alter verdeutlichen die Altersstruktur eines Wahlkreises. Diese ist vom Bundeswahlleiter gegeben. Fünf Altersgruppen und deren Bevölkerungsanteile werden geprüft. Die Einteilung lautet: Gruppe *age1*: unter 18 Jahre, Gruppe *age2*: 18 - 25 Jahre, Gruppe *age3*: 25 – 35 Jahre, Gruppe *age4*: 35 - 60 Jahre und Gruppe *age5*: über 60 Jahre.

Bildung sind Regressoren, deren Datenmengen aus den Anteilen der Schulabgänger mit unterschiedlichen Schulabschlüssen besteht. Damit kann ein eventueller Einfluss des Bildungsniveaus auf die gewählte Partei im Hinblick auf ihre Bildungspolitik nachgewiesen werden. Die Variablen sind in Prozent direkt vom Bundeswahlleiter gegeben. Für das Wahljahr 2002 wurden Zahlen aus 1999, für 2005 Zahlen aus 2002, für 2009 Zahlen aus 2007 und für 2013 Zahlen aus 2011 einbezogen. Bis auf Daten aus 2002 wurden für die restlichen Repräsentativstatistiken Zahlen von vor zwei Jahren genutzt. Eine mögliche Verfälschung der Regression wird nicht angenommen. Die vom Bundeswahlleiter gegebenen Prozentzahlen der Schulabgänger ohne Abschluss (Gruppe *edu1*), mit Hautschulabschluss (Gruppe *edu2*), mit Realschulabschluss (Gruppe *edu3*) und mit allgemeiner Hochschulreife (Gruppe *edu4*) werden ebenfalls betrachtet.

Vom Bundeswahlleiter werden auch die Anteile der sozialversicherungspflichtigen Beschäftigten in den unterschiedlichen Sektoren veröffentlicht. Neben „Land-, Forstwirtschaft und Fischerei“ (Gruppe *sec1*) publiziert der Bundeswahlleiter unter anderem auch Zahlen zum „Produzierenden Gewerbe“ (Gruppe *sec2*), „Handel, Gastgewerbe und

⁴ Bundesagentur für Arbeit

Verkehr“ beziehungsweise „Handel, Verkehr und Nachrichtendienst“ (Gruppe *sec3*) und „öffentliche Dienstleistungen“ und „übrige Dienstleistungen“ (beide ergeben zusammen Gruppe *sec4*).

2.3 Spezifikation

Soeben wurden die zu untersuchten Variablen beschrieben. Es bestünde die Möglichkeit, eine Paneldatenanalyse durchzuführen. Dabei werden N Individuen zu T Zeitpunkten beobachtet. Diese Voraussetzung wird jedoch nicht erfüllt, da sich die Bevölkerung der untersuchten Wahlkreise ändert. Insbesondere sind durch Bevölkerungswanderung zwischen den Wahlkreisen potentielle Verzerrungen möglich.

Bei einer gepoolten Regression werden in unterschiedlichen Zeitpunkten verschiedene Individuen beobachtet. Die erhaltenen Informationen in verschiedenen t werden zusammengefasst und regressiert. Jedes Jahr werden dieselben Wahlkreise untersucht. Zudem haben Wahlstatistiken anonymen Charakter, sodass dieselben Individuen in einem Wahlkreis untersucht werden können.

Die Methode der multiplen linearen Regressionsanalyse stellt weniger Anforderungen an den zu untersuchten Daten. Sie eignet sich besonders zur Untersuchung von linearen Einflüssen mehrerer Faktoren auf eine Variable. Kennzeichnend für das lineare Regressionsmodell ist die Additivität und Linearität seiner systematischen Komponenten $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$. Die Existenz einer Linearität in den Parametern ist Voraussetzung für die Anwendung des linearen Modells⁵. Das allgemeine multiple Regressionsmodell hat eine ähnliche Terminologie wie das einfache Regressionsmodell – besteht jedoch aus vielen Regressionsparametern:

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_k x_k + e_i. \quad (2.1)$$

Im obigen wahren Modell sind $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ unbekannte Parameter, y_i ist die abhängige Variable und e_i sind „unobserved random errors“ (URE). Diese können nicht ermittelt werden. Mit K als Anzahl der unabhängigen Variablen sind die erhobenen Strukturdaten,

⁵ Vgl. Fahrmeier, 2010, S. 508

wobei $k = 1, 2, \dots, m$. Die Anzahl der Beobachtungen ($i = 1, 2, \dots, n$) entsprechen der Anzahl der Wahlkreise. Die unabhängigen Variablen, die aus den gegebenen Strukturdaten entnommen werden, sind auf Wahlkreisebene erhoben und werden im multiplen Regressionsmodell mit den Wahlergebnissen hinsichtlich der Erst- und Zweitstimme des jeweiligen Wahlkreises regressiert. Analog lautet die Formel für das zu erforschende Modell:

$$\text{Stimmenanteil}_i = \beta_0 + \beta_1 \text{sex} + \beta_2 \text{mig} + \beta_3 \text{une} + \beta_4 \text{age} + \dots + \beta_5 \text{sec}_4 + e_i \quad (2.2)$$

(vgl. Wooldridge, 2013 S. 65). Die Regressionskoeffizienten β_k werden mittels der gegebenen X - und Y -Daten geschätzt. Das geschätzte Modell, das mittels SPSS ermittelt wird, hat die Form:

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + \dots + b_k x_k \quad (2.3)$$

Diese Funktion bietet für Beobachtungen x_i die geschätzten Werte \hat{y} . Die Lage der Regressionsgerade wird durch das konstante Glied (Interzept) b_0 und die geschätzten Regressionskoeffizienten b_1 bis einschließlich b_K bestimmt. Hierbei ist b_0 der Schätzer für β_0 , b_1 der Schätzer für β_1 , usw.. Die Regressionsgerade zeigt die Richtung des linearen Zusammenhangs zwischen den Variablen. Nach dem Prinzip der kleinsten Quadrate (KQ-Methode) sind die *geschätzten* Parameter β_k so zu bestimmen, dass die Summe der quadrierten Abweichungen zwischen den beobachteten Y und den erwarteten \hat{Y} minimal sind. In Matrixform lautet das Modell:

$$Y = X\beta + e \quad (2.4)$$

$$\begin{pmatrix} y_1 \\ \vdots \\ y_I \end{pmatrix} = \begin{pmatrix} x_{11} & \dots & x_{1K} \\ \vdots & \ddots & \vdots \\ x_{I1} & \dots & x_{IK} \end{pmatrix} \cdot \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_I \end{pmatrix} + \begin{pmatrix} e_1 \\ \vdots \\ e_I \end{pmatrix}$$

Die KQ-Schätzung für den β -Parameter wird widerspiegelt in der Formel:

$$b = (X'X)^{-1}X'y \quad (2.5)$$

Der Vektor der angepassten Werte \hat{y} lässt sich aus dem Produkt der Matrix X und dem Vektor b berechnen.

KQS besitzt fünf statistische Eigenschaften (E1 bis E5):

E1: Als eine Funktion von y ist die KQS eine Zufallsvariable.

E2: Die KQS ist in y linear, d.h. $b = Ay$ mit $A = (X'X)^{-1}X'$.

E3: Außerdem ist die KQS erwartungstreu: $E[b] = \beta$.

E4: Kovarianzmatrix $\sum_b = Cov[b] = E[(b - \beta) \cdot (b - \beta)'] = \sigma^2(X'X)^{-1}$

Geschätzte Kovarianzmatrix $\widehat{Cov}[b] = E[(b - \widehat{\beta})(b - \widehat{\beta})'] = \widehat{\sigma}^2(X'X)^{-1}$

$$\Rightarrow E[\widehat{\sigma}^2] = \sigma^2$$

E5: Nach dem Gauß-Markov-Theorem ist KQS effizient (BLUE = Best Linear Unbiased Estimator), also der beste lineare, unverzerzte Schätzer.

Die Residualgröße \hat{e}_i , lässt sich aus der Differenz von y_i und \hat{y}_i errechnen

$$\hat{e}_i = y_i - \hat{y}_i \quad (2.6)$$

Sie kann für einzelne Beobachtungen auch in folgender Form umschrieben werden

$$\sum_{i=1}^n (y_i - b_0 - b_1x_{1i} + b_2x_{2i} + b_3x_{3i} + b_4x_{4i} + b_5x_{5i})^2 \quad (2.7)$$

Die KQ-Methode, die auch Ordinary Least Squares Methode (OLS) genannt wird, minimiert die Summe der quadrierten Residuen $\sum \hat{e}_i^2$. Für jede Beobachtung existiert ein Residuum.

Wenn $\hat{e}_i > 0$, bedeutet dies, dass \hat{y}_i kleiner als y_i ist, so wäre y_i unterschätzt.

Andersherum liefert $\hat{e}_i < 0$ mit \hat{y}_i größer als y_i ein überschätztes y_i . Die Zielfunktion der untersuchten multiplen Regression lautet:

$$\min_{b_0, b_1, \dots, b_k} \sum_{k=1}^K [e_k]^2 = \min_{b_0, b_1, \dots, b_k} \sum_{k=1}^K [y_k - (b_0 + b_1x_{1i} + b_2x_{2i} + b_3x_{3i} + \dots + b_kx_{ki})]^2 \quad (2.8)$$

mit e_i als Residualgröße, x_{ik} als unabhängigen Variablen, $b_0, b_1, b_2, \dots, b_k$ als Regressionskoeffizienten. Damit die oben genannte Minimierung eine eindeutige Lösung besitzt, sollte der Umfang der Daten mindestens so groß sein wie die Anzahl der unbekannten Parameter K . Je größer der Datenumfang n im Bezug zum k , desto kleiner ist der Schätzfehler (Fahrmeir, 2010 S. 496). Dies ist bei dem vorliegenden Datensatz erfüllt, da

es für jeden der 299 Wahlkreise eine Beobachtung, also $n = 299$ Beobachtungen gibt. Zudem müssen wenigstens die abhängigen Variablen metrisch skaliert sein, während unabhängige Variablen durch Dummyvariablen nominalskaliert sein können. Außer diesen allgemeinen Regeln sind nach Auer (2011) weitere Annahmen für das multiple Regressionsmodell gegeben. Nun werden die Prämissen genannt und die Techniken zur Prüfung dieser bündig erläutert.

Annahme 1: In Gleichung (1) sind keine irrelevanten exogenen Variablen einbezogen. Zudem fehlen keine relevanten Prädikatoren, sodass die korrekte Variablenwahl getroffen ist. Diese Variablen werden aus den gegebenen Strukturdaten entnommen. Prädikatoren wie Migration und Unemployment sind nicht gruppiert und nach einer Korrelationsuntersuchung für die Regression zugelassen wurden. Wie zuvor bereits beschrieben, sind deren Einflüsse außerdem auch theoretisch erklärbar. Bei normalverteilten Daten und einem linearen Zusammenhang zwischen den Variablen wird der Pearson-Korrelationskoeffizient angewandt, sonst sollte der Spearman-Korrelationskoeffizient betrachtet werden.

Annahme 2: Der wahre Zusammenhang zwischen den erklärten und den erklärenden Variablen ist *linear*. Diese Beziehung zwischen den Variablen wurde bereits im Punkt 2.1 theoretisch angerissen. Die Annahme wird im Kapitel 3.1 mit Streudiagrammen überprüft. Das Streudiagramm, das bei der Regression angegeben werden kann, zeigt die standardisierten Residuen und standardisierte Y -Werte. Der optische Test ob die Streuung kurvilinear ist, dient als Indikator für lineare Zusammenhänge.

Annahme 3: Für alle n Beobachtungen werden konstante Parameter angenommen.

Annahme 4: Der Erwartungswert der Residuen ist Null: $E(e_i) = 0$ für $i = 1, 2, \dots, n$.

Annahme 5: Die Varianz der Residuen bleibt für alle Beobachtungen gleich, d.h. $Var(e_i) = \sigma^2$ (Homoskedastizität). Wenn X -Werte größer werden, streuen Y -Werte nicht stärker, sondern bleiben gleich bezüglich der Streuung. Sind die Varianzen unterschiedlich, wären die Daten heteroskedastisch. Die Homoskedastizitätsprüfung kann graphisch mit Streudiagrammen oder Boxplots approximativ durchgeführt werden. Bei dieser Studie wird wie bei Linearitätsprüfung mit Streudiagrammen gearbeitet. In diesen sollten keine Muster erkennbar sein, da dies eine zufällige Streuung vermuten lässt. Wenn eine bestimmte Systematik zu sehen ist, ist die Annahme der Varianzhomogenität verletzt. Je nachdem wie schwer, wirkt sich dies auf die Regression aus. Im besten Fall liegen sämtliche Residuen

horizontal verteilt und nehmen weder zu noch ab. Je weniger Ausreißer existieren, desto mehr spricht der optische Test gegen Heteroskedastizität. Die Prüfung auf Homoskedastizität erfolgt im Textabschnitt 3.1.

Annahme 6: Außerdem werden Residuen als normalverteilt angenommen. Ob Residuen normalverteilt sind, ist mit Histogrammen erkennbar. Bei einem zusätzlichen P-P-Diagramm, welches von SPSS geplottet wird, werden die beobachteten gegen die erwarteten standardisierten Residuen geplottet. Die beobachteten liegen bei Linearität auf der diagonalen Linie. Der optische Test auf Normalverteilung der Residuen befindet sich ebenfalls im Kapitel 3.1.

Annahme 7: Ferner werden exogene Variablen deterministisch angenommen – gleichbedeutend mit kontrollierbar. Somit sind sie keine Zufallsvariablen.

Annahme 8: Zwischen den anhängigen Variablen muss perfekte Multikollinearität ausgeschlossen sein. Diese Annahme wird später mit einer Pearson-Korrelationsanalyse geprüft. Sie analysiert die Korrelation und dessen Signifikanz zwischen zwei Variablen. Ein Pearson-Korrelationskoeffizient in Wert von 1 oder -1 bedeutet perfekte Korrelation. Die Vermutung, dass die Variablengruppen *age1, age2, age3, age4* und *age5; adu1, edu2, edu3* und *edu4; sec1, sec2, sec3* und *sec4* miteinander korrelieren, ist begründbar. Jeder Regressor kann durch die Differenz zu anderen Variablen derselben Gruppe erklärt werden. Eine lineare Regressionsanalyse dieser Regressoren hilft den Zusammenhang aufzudecken. Dabei wird eine der Variablen als abhängige Variable behandelt und die anderen als unabhängige. Das Bestimmtheitsmaß erlaubt eine Aussage über die Einflüsse mehrerer Regressoren auf einen Regressor. Die Ergebnisse dieser Untersuchung werden wie andere Annahmen in 3.1 erläutert.

2.4 Regression und Prüfung

Es ist sinnvoll eine Rückwärtsregression via SPSS durchzuführen, da vermutet wird, dass sämtliche Prädiktoren eine lineare Beziehung zur abhängigen Variablen haben. Nach und nach werden unabhängige Variablen sukzessiv aus dem Modell entnommen, indem die Prämisse mit der geringsten Teilkorrelation ausgeschlossen wird. Anschließend wird die nächste Variable entnommen usw. Dieser Vorgang läuft so lange, bis alle Variablen, die das Ausschlusskriterium erfüllen, aus dem Modell entfernt worden sind. Damit kann im Zusammenhang mit den auserwählten Prämissen erreicht werden, dass das bestmögliche Modell gefunden wird und gleichzeitig so viele relevante, unabhängige Variablen einbezogen werden, wie möglich. Für jedes Wahljahr werden die Wahlergebnisse mit den 16 Prädiktoren regressiert.

Die durch SPSS angegebenen Regressionskoeffizienten drücken den Einfluss der Prädiktoren aus. Je nach Vorzeichen und Höhe der Regressionskoeffizienten reduzieren oder erhöhen sie die Ergebnisse der Bundestagswahlen. Diese unstandardisierten Regressionskoeffizienten eignen sich, um gleiche Regressionsmodelle zu vergleichen. Bei dieser Studie soll jedoch auch auf die standardisierten Regressionskoeffizienten geachtet werden. Bei dieser linearen Regression sind sowohl abhängige als auch unabhängige Variablen standardisiert. Dies bedeutet, dass der Mittelwert bei Null und die Varianz bei Eins liegen, d.h. $\mu = 0; \sigma^2 = 1$. In der Theorie ist dieser Beta-Wert aus den Regressionskoeffizienten errechenbar:

$$\tilde{b}_i = b_i \cdot \frac{s_{xi}}{s_y} \quad (2.9)$$

Durch eine Standardisierung der Variable können die Einflüsse der einzelnen Prädiktoren auf die abhängigen Variablen geschätzt werden. Die standardisierten Regressionskoeffizienten liegen auf dem Intervall $[-1, 1]$ und sind unabhängig von den ursprünglichen Maßeinheiten der unabhängigen Variablen. Die Richtung des Einflusses wird durch das Vorzeichen und die Stärke durch den Betrag ausgedrückt. Die Variable mit dem höchsten Beta-Betrag hat den größten Einfluss. Je mehr sich der Regressionskoeffizient Null nähert, desto unwahrscheinlicher ist die Wahrscheinlichkeit auf einen linearen Zusammenhang. Diese Technik eignet sich besonders, um verschiedene Modelle zu

vergleichen. Schließlich wird davon ausgegangen, dass unterschiedliche multivariate lineare Regressionsgleichungen als Ergebnisse geplottet werden. Insbesondere die theoretische Vermutung, dass unterschiedliche Wahlkreiseigenschaften unterschiedlich starke Einflüsse auf Wahlbeteiligung bzw. Wahlergebnisse haben, unterstützt diese Vorhersage. In einer Regressionsgleichung mit standardisierten Regressionskoeffizienten sind alle Mittelwerte der Variablen Null. Daher existiert kein konstanter Term wie im unstandardisierten Modell. Da die Interpretationsfähigkeit der standardisierten Regressionskoeffizienten vielen Literaturquellen zu Folge umstritten ist, werden sie bei dieser Studie nur genutzt, um die Richtung und Stärke des Einflusses zu erkennen. Zur Interpretation werden die unstandardisierten Koeffizienten verwendet.

Nun werden zuerst die Kennwerte zur globalen Prüfung des Regressionsmodells und danach zur Prüfung der einzelnen Regressionskoeffizienten angegeben. Auch die Interpretationsmöglichkeiten werden dargelegt. Es wird erörtert, inwiefern die Kennwerte eine Aussage über die Erklärungsfähigkeit des Modells treffen können. Um das gesamte Modell zu prüfen, wird mit dem adjustierten Bestimmtheitsmaß, dem F -Test und dem Standardfehler gearbeitet. Die Kontrolle der Regressoren erfolgt dann mittels Betrachtung der t -Werte, deren Signifikanzniveaus, Konfidenzintervalls, Toleranz und Varianzinflationsindikatoren.

Globale Prüfung

Das Bestimmtheitsmaß R^2 welches berechnet oder vom Programm geplottet werden kann, gibt Aussage darüber, wie gut das Modell bzw. die Regressionsgerade die Beziehung zwischen den unabhängigen und den abhängigen Variablen beschreibt. Es ist damit ein Gütemaß der linearen Regression. Aus dem Bestimmtheitsmaß kann der Anteil der Variation in Y abgelesen werden, der mit Hilfe des multiplen linearen Regressionsmodells durch die Variation in X erklärt wird.

$$R^2 = \frac{\sum_{k=1}^K (\hat{y}_k - \bar{y})^2}{\sum_{k=1}^K (y_k - \bar{y})^2} = \frac{SSR}{SST} = \frac{\text{erklärte Streuung}}{\text{Gesamtstreuung}} \quad (2.10)$$

$$\begin{aligned}
&= 1 - \frac{\sum_{k=1}^K (y_k - \hat{y}_k)^2}{\sum_{k=1}^K (y_k - \bar{y})^2} = 1 - \frac{\sum_{k=1}^K e_k^2}{\sum_{k=1}^K (y_k - \bar{y})^2} = 1 - \frac{SSE}{SST} \\
&= 1 - \frac{\text{nicht erklärte Streuung}}{\text{Gesamtstreuung}}
\end{aligned} \tag{2.11}$$

Zum Vergleich verschiedener Modelle eignet sich das R^2 nicht, da es durch die Aufnahme weiterer Variablen automatisch zunimmt. Wenn kein Zusammenhang zwischen X und Y besteht, bleibt R^2 zwar unverändert, steigt dafür bei einem minimalen Zusammenhang. Es sei darauf hingewiesen, dass für das Jahr 2009 Strukturdaten nicht vollständig gegeben sind, sodass die fehlenden Variablen nicht in das Modell einfließen können. Damit unterscheidet sich das Modell vom Wahljahr 2009 von den restlichen. Wäre dem nicht so, könnte R^2 genutzt werden. Zur Beurteilung der Güte von unterschiedlichen linearen Regressionsmodellen eignet sich das angepasste (auch korrigierte, adjustierte oder bereinigte) R^2 (\bar{R}^2) besser, da es auch die Freiheitsgrade der Regression berücksichtigt. \bar{R}^2 berücksichtigt hingegen die Anzahl an unabhängigen Variablen und verhindert, dass durch die Hinzunahme eines Prädiktors der Anteil der erklärten Variation nur dadurch steigt. Auf diese Weise verhindert es auch, dass Variablen wahllos in das Modell eingebracht werden. So kann die fälschliche Interpretation eines vermeintlich guten Regressionsmodells unterbunden werden. Auch \bar{R}^2 liegt zwischen Null und Eins und deutet bei einer Annäherung an Eins auf ein gutes Modell hin⁶.

$$\bar{R}^2 = R^2 - \frac{J \cdot (1 - R^2)}{K - J - 1} = 1 - \frac{SSE / (I - K)}{SST / (I - 1)} = 1 - \left[\frac{I - 1}{I - K} \cdot (1 - R^2) \right] \tag{2.12}$$

K entspricht der Anzahl der Beobachtungswerte und J die der Regressoren. An dieser Gleichung lässt sich unschwer erkennen, dass durch eine Zunahme der Regressoren \bar{R}^2 nicht zwangsläufig steigt. Da die Anzahl der Regressoren k auf der einen Seite dazu führt, dass R^2 sich erhöht, jedoch auf der anderen Seite \bar{R}^2 wegen dem Term $\frac{I-1}{I-K}$ sich auch verkleinern kann. Abhängig davon, welcher Effekt stärker wirkt, sollte die entsprechende Variable in das Modell einbezogen werden. Durch die „Strafung“ der Regressorenzunahme wird ein kleineres (oder maximal gleiches) \bar{R}^2 als R^2 erwartet.

⁶ vgl. Backhaus, 2011 S. 74 – 76

Ein weiterer Indikator, der bei der Modellprüfung wichtig ist, ist die F -Statistik. Diese prüft die Signifikanz des Bestimmtheitsmaßes unter der Nullhypothese H_0 : *Es besteht kein Zusammenhang zwischen der abhängigen Variable y und der unabhängigen Variable x* – gleichbedeutend mit $H_0: \beta_1 = \beta_2 = \dots = \beta_j = 0$. Das Modell wäre damit nutzlos. Üblich wird der empirische Wert der F -Statistik mit der Formel:

$$F_{emp} = \frac{\sum_{i=1}^N (\hat{y}_i - \bar{y})^2 / k}{\sum_{i=1}^N (y_i - \hat{y}_i)^2 / (N - k - 1)} = \frac{R^2 / k}{(1 - R^2) / (N - k - 1)} \quad (2.13)$$

errechnet und schließlich mit einem kritischen Wert verglichen. Diese F -Verteilung besitzt k Zählerfreiheitsgraden und $N - k - 1$ Nennerfreiheitsgraden. Für die Regressionsanalyse ist das gewählte Konfidenzintervall 0,95. Die Irrtumswahrscheinlichkeit liegt damit bei $\alpha = 1 - 0,95 = 0,05$ und entspricht einer fünf prozentigen Fehlerwahrscheinlichkeit des Tests. Auch in dieser Arbeit wird dieser Wert genutzt. Ist der F -Wert nahe Null, kann die Nullhypothese verworfen werden. Überschreitet dieser Wert hingegen den kritischen Wert (auch theoretischer Wert genannt) oder weicht stark von Null ab, muss die Nullhypothese verworfen werden. In dieser Situation kann ein Zusammenhang zwischen den Variablen deduziert werden.⁷ Der theoretische F -Wert kann aus Tabellen (vgl. Andreß, 2000) entnommen werden. Der Zusammenhang ist signifikant, wenn die Nullhypothese verworfen wird – also wenn $F_{emp} > F_{theo}$. Dabei kann $F_{emp} < F_{theo}$ jedoch nicht als „kein Zusammenhang zwischen den Variablen“ interpretiert werden. Der Einfluss einer Variablen ist eventuell durch andere Prädikatoren überlagert oder wegen Auslassung relevanter Einflussgrößen nicht erfassbar.⁸ Dieser Undeutlichkeit zufolge ist die Annahme der Nullhypothese keine Gewährleistung für die Nicht-Existenz eines Zusammenhangs. Eine Ablehnung von H_0 ist dagegen ein eindeutiger Hinweis auf Signifikanz. Bei der gewählten Methode der Rückwärtsregression wird die F -Wahrscheinlichkeit verwendet. Der typische Aufnahmewert von 0,05 und Ausschlusswert von 0,10 werden auch bei dieser Studie genutzt, um ein geeignetes Modell zu finden. Sollte die Nullhypothese gültig sein, kann mit einem F -Wert von Null gerechnet werden. Je mehr dieser eindeutig von Null abweicht, desto eher erfolgt eine Ablehnung der Nullhypothese. Es kann davon ausgegangen werden,

⁷ Vgl. Backhaus, 2011 S. 77

⁸ Vgl. Backhaus, 2011 S. 80

dass bei einem hohen Bestimmtheitsmaß der F -Wert dementsprechend hoch ist. Aus dem Grund, dass der F -Wert mit dem Zusammenhang zwischen abhängigen und unabhängigen Variablen steigt. So kann ein Zusammenhang der Grundgesamtheit interpretiert werden. Dank der relativ großen Grundgesamtheit, kann mit einem ebenfalls starken F -Wert gerechnet werden. Im Allgemeinen gilt, je größer die Stichprobe, desto geringer die Unsicherheit darüber, dass die Parameter ungleich Null sind. Der F -Wert und sein Signifikanzniveau werden in SPSS ausgegeben, sodass ein Vergleich zwischen dem empirischen und dem theoretischen nicht mehr nötig ist, um das Signifikanzniveau herauszufinden.

Der Standardfehler der Schätzung ist der „mittlere Fehler bei Verwendung der Regressionsfunktion zur Schätzung der abhängigen Variablen“ y .⁹ Die Berechnung erfolgt mit der Formel:

$$\sigma = \sqrt{\frac{\sum_{k=1} e_k^2}{K - I - 1}} \quad (2.14)$$

Diese Formel wurde aus der Formel zur Berechnung der Varianz abgeleitet. Bei der Interpretation des Standardfehlers wird auch der Mittelwert \bar{y} in Betracht gezogen und je nach \bar{y} bewertet. Der Standardfehler der Schätzung hat eine ähnliche Aussagekraft wie das korrigierte Bestimmtheitsmaß und gibt eine Aussage über die „Anpassungsgüte der Regressionsgleichung“. Werden die 299 Wahlkreise als Grundgesamtheit betrachtet, kann der berechnete σ als wahre Abweichung angesehen werden. Genauer genommen, ist die Wahlstatistik jedoch eine repräsentative Statistik, wo Zahlen der Wähler hochgerechnet werden. Somit ist der Standardfehler im eigentlichen Sinne kein wahrer Wert. Diese Anmerkung sollte beachtet werden, obwohl die Wahlkreise durchaus als eine Art Grundgesamtheit angesehen werden können, da wohl keine genaueren Statistiken möglich sind.

Alle einbezogenen unabhängigen Variablen sollten einzeln geprüft werden. Dadurch wird erkennbar, welche tatsächlich in das Modell gehören und welche nicht. Alle Variablen

⁹ vgl. Backhaus, 2011 S. 80

werden durchgeprüft. Bei einer multiplen Regression wie diese, gilt für die Stichprobenverteilung von b_i :

$$b_i \sim N(\beta_i, \sigma_\varepsilon^2 (X'X)_{ii}^{-1}) \quad (2.15)$$

Mit $(X'X)_{ii}^{-1}$ als das i -te Hauptdiagonalelement der Matrix $(X'X)^{-1}$. Das geschätzte σ wird ebenfalls von SPSS ausgegeben. Dieser errechnet sich sonst aus der vorherigen Formel, sodass:

$$\hat{\sigma}_{bi} := \sqrt{\hat{\sigma}_\varepsilon^2 (X'X)_{ii}^{-1}} \quad (2.16)$$

Prüfung der Regressionskoeffizienten

Die Regressionskoeffizienten werden mittels t -Test geprüft, um zu erfahren welche unabhängigen Variablen für das Modell relevant sind. Der t -Wert berechnet sich aus dem Quotienten von $(b_i - \beta_i)$ und dem Standardfehler σ_{bi} .

$$t_{emp} = \frac{(b_i - \beta_i)}{\hat{\sigma}_{bi}} \quad (2.17)$$

Unter der Nullhypothese H_0 : *Die t – Statistik ist im Mittel Null*, kurz: $\beta_i = 0$, liegen die empirische t -Werte um Null. Weichen diese jedoch stark ab, wird die Nullhypothese verworfen. Somit ist der wahre Parameter als ungleich Null zu interpretieren. Die Verwerfung der Nullhypothese bedeutet, dass zwischen y_i und x_i ein Zusammenhang besteht. Auf das Signifikanzniveau des Tests sollte ebenfalls geachtet werden. Der Test ist umso besser, je kleiner der Alpha-Fehler. Bei dieser Studie wird ein zweiseitiger t -Test durchgeführt.

Das Konfidenzintervall um einen Regressor ist der sogenannte Vertrauensbereich. Es sollte nicht zu groß ausfallen, da dies auf eine hohe Unsicherheit des Prädiktors deutet. Bei dieser Studie wird ein Konfidenzintervall von 95 Prozent gewählt. Die Lage des wahren Parameters liegt also in 95 Prozent der Fälle in diesem Intervall. Formal kann das Intervall mit:

$$KI_{0,95} := [b_i - t \cdot \hat{\sigma}_{bi} \leq \beta_i \leq b_i + t \cdot \hat{\sigma}_{bi}] \quad (2.18)$$

ermittelt werden.

Die Toleranz ist die Differenz von Eins und dem multivariaten Korrelationskoeffizienten. Ist die Toleranz gering (kleiner als 0,1), deutet sie auf Multikollinearität zwischen den unabhängigen Variablen. Bei einem Toleranzwert unter 0,01 kann sicher eine Multikollinearität interpretiert werden. Der Kehrwert der Toleranz ist der Varianzinflationsfaktor (VIF). Dieser kann ebenfalls bei Spekulation auf Multikollinearität angewendet werden. Ein VIF-Wert ab 10 lässt multikollineare Variablen vermuten, während ein Wert ab 100 zuverlässig Multikollinearität signalisiert. Es kann bereits vorausgesagt werden, dass Variablen, die zusammenhängend sind, kollinear sein können. Beispielsweise werden gruppierte Daten wie *age1*, *age2*, *age3*, *age4* und *age5*, wenn alle Variablen in ein Modell eingehen, Kollinearität verursachen. Grund dafür ist, dass die Prädiktoren Anteile der Altersgruppen angeben. Die Summe dieser Prädiktoren beträgt eins. Die Variable *age1* kann etwa durch die Differenz von Eins und die Summe der restlichen Variablen *age2*, *age3*, *age4* und *age5* berechnet werden. Dies gilt auch für Prädiktoren, die die Anteile der Schulabsolventen und die Anteile der sozialversicherungspflichtigen Beschäftigten in den Sektoren beschreiben. Regressionskoeffizienten mit der stärksten Multikollinearität können aus der Regressionsgleichung eliminiert werden. Die Folge dieser Beseitigung ist ein möglicherweise falsch spezifiziertes Modell. Daraus entstünde systematisch verzerrte Schätzer.

3 Regression

Vor der Realisation der Regressionsanalyse, sollten die Voraussetzungen für die Durchführung einer Regression kontrolliert werden. Prämissen bezüglich Residualgrößen (Normalverteilung, Homoskedastizität, Autokorrelation), unabhängige Variablen (perfekte Multikollinearität) und abhängige Variablen (Linearer Zusammenhang zwischen X und Y) werden getestet. Im Anschluss werden die Regressionen und deren Ergebnisse ausgearbeitet.

3.1 Prüfung der Voraussetzungen

Um die erklärenden Variablen auf Multikollinearität zu prüfen, wird eine Pearson-Analyse durchgeführt. Es wird keine Spearman-Analyse durchgeführt, da bei der Studie keine ordinalskalierten Daten auftreten. Zudem ist die „Stichprobe“ sehr hoch (größer als 20). Korrelationskoeffizienten der Pearson-Korrelation, die höher sind als 0,5, sind in den Tabelle 3.1 und Tabelle 3.2 grau markiert. Die meisten der Korrelationskoeffizienten sind geringer als 0,7. Dies bedeutet, dass keine hohe Korrelation zwischen den betroffenen unabhängigen Variablen besteht. Die Korrelationskoeffizienten, die mit einem Sternchen markiert sind, sind bei Niveau 0,05 signifikant, die mit zwei Sternchen bei Niveau 0,01. Der Anteil der Altersgruppe der unter Achtzehnjährigen (*age1*) weist von häufig hohen Korrelationen mit anderen Variablen auf. Darunter fallen die Variablen *age1*, *edu2*, *edu4*, *sec2* und *sec4*. Die Korrelation dieser Daten sind deutlich signifikant. Ebenso oft tritt auch bei *edu4* (Anteil der Schulabsolventen mit Fachhochschulreife) ein mittlerer Zusammenhang auf, welche höher sind als 0,5. Sie korreliert mittelstark negativ mit den unabhängigen Variablen *age1*, *edu2*, *sec* und positiv mit *sec4*. Eine sehr negative, hohe und signifikante Korrelation wird zwischen *sec2* (Anteil der sozialversicherungspflichtigen im Sektor Produzierendes Gewerbe) und *sec4* (Anteil der sozialversicherungspflichtigen im Sektor Dienstleistungen) beobachtet. Diese sollten möglichst nicht gemeinsam in einem Modell auftreten. Mittelhohe, signifikante Koeffizienten sind umrahmt. Diese kommen bei *une* (Arbeitslosenquote), *edu2* (Anteil der Schulabsolventen mit Hauptschulabschluss) und *edu3* (Anteil der Schulabsolventen mit Realschulabschluss) vor.

Die lineare Regressionsanalyse für *age1*, *edu1* und *sec1* mit Variablen aus derselben Gruppe zeigen, dass ein beinahe perfekter Zusammenhang zwischen ihnen existiert (Siehe Tabelle 6.2). Multikollinearität kann auf die Gefahr einer Falschspezifikation, durch Weglassen einer Variablen absorbiert werden. Einzig der Anteil der Schulabgänger (*edu1*) lässt sich weniger gut durch die verwandten Daten erklären. Erklärung dafür ist eine Gruppe, die nicht in die Untersuchung einbezogen wird. Über einen kleinen Anteil der Schulabsolventen existiert keine genaue Information über ihre Abschlüsse.

Modellübersichten				
Modell	R	R-Quadrat	Angepasstes R-Quadrat	Standardfehler der Schätzung
age1	1,000 ^a	,999	,999	,068343
edu1	,884 ^a	,782	,781	1,202787
sec1	,999 ^a	,998	,998	,057797

Tabelle 3.1: Test auf Multikollinearität zwischen bestimmten Gruppen unabhängiger Variablen durch lineare Regression

Der lineare Zusammenhang zwischen den abhängigen und den unabhängigen Variablen kann anhand von Streudiagrammen geprüft werden. Fallen Daten auf, indem sie U-förmig gestreut sind oder wie ein umgekehrtes „U“ aussehen, sollten diese genauer untersucht werden. Die Analyse verlangt hohe Kapazitäten. Daher werden nicht alle Streudiagrammen zwischen den einzelnen erklärten und erklärenden Variablen dargestellt. Im Folgenden wird repräsentativ der Zusammenhang zwischen Zeitstimmenanteile der CDU/CSU und dem Anteil der Schulabgänger mit Hauptschulabschluss im Jahr 2002 dargestellt. Anhand der Streuung lässt sich eine lineare Beziehung erraten. Die Untersuchung zeigt keine Beziehungen, die eine lineare Regressionsanalyse gefährdet.

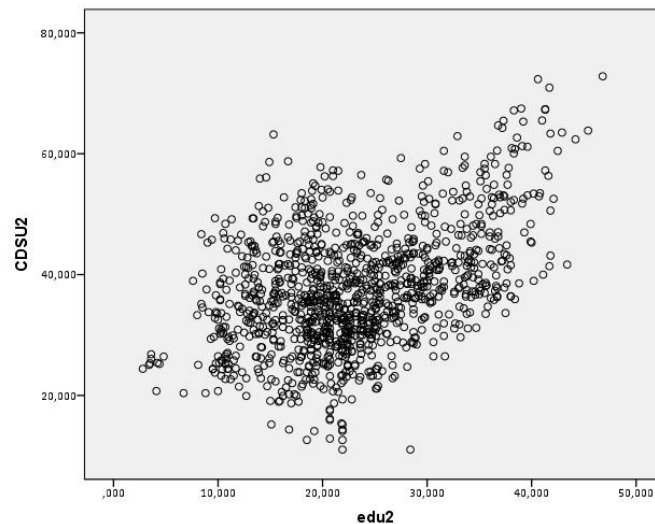


Abbildung 3.1: Streudiagramm zur Untersuchung der linearen Beziehung zwischen *CDU/CSU* und *edu2*.

Nun werden die Voraussetzungen an den Residuen kontrolliert. Dazu werden sie auf Normalverteilung und Homoskedastizität geprüft. Die optischen Prüfungen der Residuen erfolgt für jedes Jahr und für sämtliche Wahlbeteiligungen, Erst- und Zweitstimmen. Aus Kapazitätsgründen werden die Resultate nicht aufgezeichnet, sondern befinden sich in der dazu gehörigen SPSS-Datei.

Residuen bei den Regressionen der Wahlbeteiligung sind stets normalverteilt. Im Jahr 2002 zeigen sich für Erststimmenanteile der FDP, der Linken, der Grünen und Zweitstimmenanteile der Linken und der Grünen schlechtere Ergebnisse bezüglich Linearität. Die Residuen der größeren Parteien können dagegen eindeutig als linear angesehen werden. In den Histogrammen weisen sie minimale Abweichungen zur Normalverteilung auf und in P-P-Diagrammen liegen diese eng an den Diagonalen. Auch die Residuen bezüglich der Regression für Erststimmenanteil der FDP, der Linken und Zweitstimmenanteil der Linken im 16. Bundestagswahl (2005) weichen leicht von einer Normalverteilung ab. Zum 17. Bundestag sind fallen die Tests ähnlich aus. Während Residuen der Regressionen für größere Parteien weiterhin sehr deutlich als normalverteilt ausgelegt werden, zeigen Residuen der Erststimmenanteile der Linken, der Grünen und Zweitstimmenanteile der Grünen wieder Abweichungen.

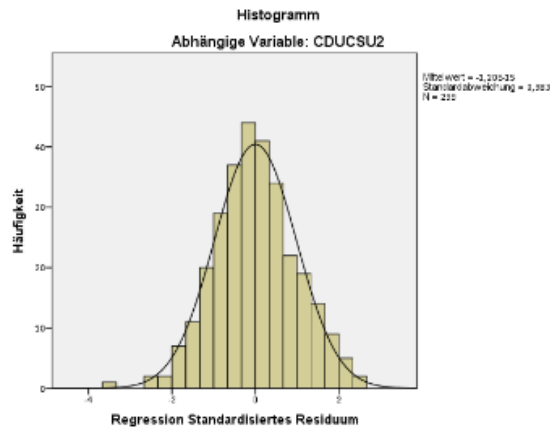


Abbildung 3.2: Histogramm zur Untersuchung der Residuen auf Normalverteilung bei Regression der Zweitstimmen CDU/CSU 2013.

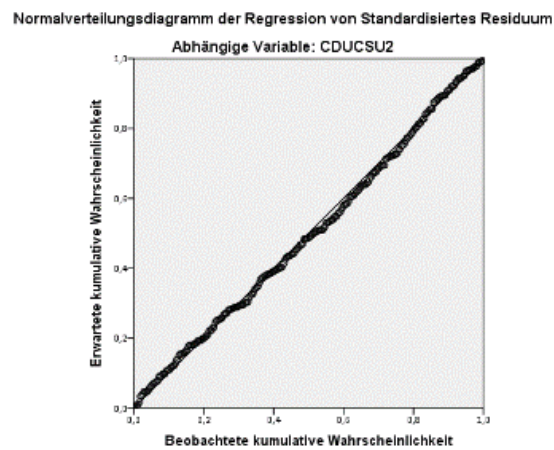


Abbildung 3.3: P-P-Diagramm zur Untersuchung der Residuen auf Normalverteilung bei Regression der Zweitstimmen CDU/CSU 2013.

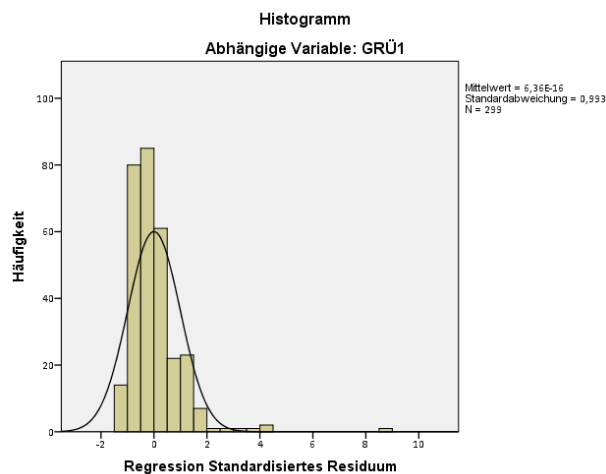


Abbildung 3.4: Histogramm zur Untersuchung der Residuen auf Normalverteilung bei Regression der Erststimmen der Grünen 2013.

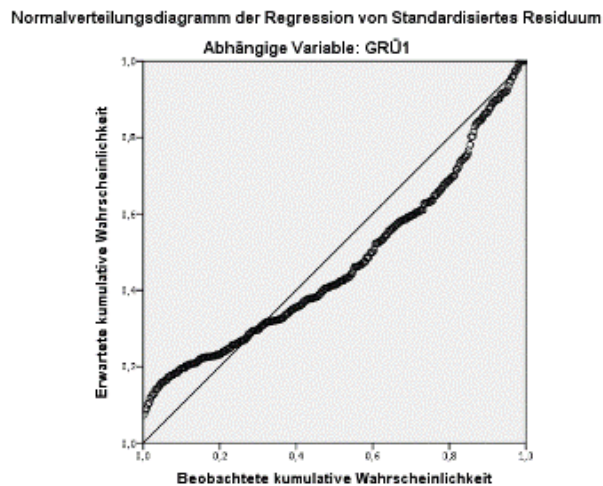


Abbildung 3.5: Histogramm zur Untersuchung der Residuen auf Normalverteilung bei Regression der Erststimmen der Grünen 2013.

Die Abbildungen zu den Verteilungen der Residuen des Wahljahres 2013 verdeutlichen nochmal, dass kleine Parteien wie die FDP und die Grünen leicht von Normalverteilungsbedingung abweichen. Die meisten der abweichenden Residuen zeigen sich bei Regression mit der Erststimme. Jedoch weichen die Werte der Histogrammbalken nicht zu extrem von einer Normalverteilung. Zusätzlich differenzieren die Werte beim Plotten der beobachteten und der geschätzten Residuen im P-P-Diagramm nicht zu sehr von der Diagonale. So kann eine Regressionsanalyse dennoch durchgeführt werden. Um die Untersuchung zu veranschaulichen werden zum einen Graphiken gezeigt, die zum einen einer Normalverteilung stark ähneln und zum anderen davon abweichen. Die Abweichungen der Residuen von der Normalverteilung sind in der Schiefe und in den Ausreißern zu erkennen.

Da die Abbildung der Homoskedatizitäts-Resultate umfangreich ist, wird auch hier repräsentativ eine Abbildung bezüglich Regression der Zweitstimmen der SPD dargestellt.



Abbildung 3.6: Streudiagramm zur Prüfung der Residuen auf Homoskedastizität bei Regression der Zweitstimmen der SPD.

Für große Parteien wie CDU/CSU und SPD deuten die Streudiagramme eher auf Homoskedastizität hin. Die Prüfung der Residuen auf Homoskedastizität via Streudiagrammen erweisen sich für Regressionen kleinerer Parteien ebenfalls als minder. In allen Wahljahren lassen sich für kleinere Parteien wie die Linke, die Grünen und FDP systematische Musterungen der Residuen in den Diagrammen erkennen. Die heterogenaussehenden Verteilung der Residuen weisen auf die Verletzung der Homogenitätsannahme. Dies beeinträchtigt jedoch nicht schwerwiegend die Regressionsanalyse. Größtenteils lassen sich für Residuen der Wahlbeteiligung, CDU/CSU, SPD und teilweise Zweitstimmen der kleinen Parteien eine zufällige Streuung nachweisen. Mit dieser letzten Prüfung der Regressionsbedingung sind alle Voraussetzungen erfüllt, sodass die Analyse durchgeführt werden kann.

3.2 Regression

Die Betrachtung der Rückwärtsregression erfolgt in drei Schritten. Zuerst müssen anhand der Kennwerte adjustiertes Bestimmtheitsmaß, F -Wert und dem Standardfehler der Regressionen die besten Modelle gefunden werden. Im zweiten Schritt wird bei den auserwählten Modellen auf mögliche Multikollinearität zwischen den unabhängigen Variablen geachtet. Sollte eine hohe Kollinearität auftreten, wird das Modell verworfen.

Diejenigen Variablen mit der höchsten Kollinearität werden im dritten Schritt ausgeschlossen, sodass ein neues Modell entsteht.

Die Güte der Anpassungen der Regressionsmodelle für die Bundestagswahlen in Wahljahren 2002, 2005, 2009 und 2013 werden untersucht und miteinander verglichen. Die Modelle mit den höchsten Bestimmtheitsmaßen, größten korrigierten Bestimmtheitsmaßen und den geringsten Standardfehler sind ausgewählt. Für die besten Modelle werden die Regressionskoeffizienten untersucht. Zuvor aber werden die Prädiktoren der ausgewählten Modelle genauer auf Multikollinearität untersucht. Wie im Theorieteil beschrieben, wird nach Multikollinearität gesucht. Sind die Kennwerte nicht im eindeutigen Bereich, wird keine Kollinearität vermutet. Es kann beobachtet werden, dass gruppierte, zusammenhängende Variablen eine sehr kleine Toleranz oder einen hohen VIF-Wert aufweisen. Von insgesamt 44 ausgewählten Regressionsmodellen sind 21 Modelle mit Kollinearität betroffen, sodass mindestens eine erklärende Variable ausgeschlossen wird. Für die bereinigten Modelle muss nun erneut eine multivariate lineare Regressionsanalyse durchgeführt werden. Die restlichen 23 Modelle, die nicht von Multikollinearität betroffen sind, werden belassen. Im Folgenden ist eine Liste der Regressionsmodelle, die wegen Multikollinearität erneut regressiert werden müssen:

2002: *WB, CDU/CSU1, CDU/CSU2, SPD1, SPD2, FDP1, LIN1, LIN2, GRÜ1, GRÜ2*.

2005: *WB, CDU/CSU1, CDU/CSU2, SPD1, SPD2, FDP1, FDP2, LIN1, GRÜ1, GRÜ2*.

2009: *WB, SPD1, SPD2, LIN1, LIN2, GRÜ1, GRÜ2*.

2013: *LIN1, LIN2*.

Beobachtungen, die zu dieser Entscheidung führen, werden im Folgenden beschrieben.¹⁰ Ausschließlich Erst- und Zweitstimme der Die Linken sind sämtliche Regressionen für das Wahljahr 2013 nicht mit Multikollinearität behaftet. In der Zeit davor sind nur Regressionen für Zweitstimmenanteil der FDP im Jahre 2002 und Zweitstimmenanteil der Die Linken multikollinearitätsfrei. Eine Neuregression wird also mit allen Variablen der 15. und 16. Bundestagswahl –bis auf *FDP2* im Jahr 2002 und *LIN2* im Jahr 2005 durchgeführt. Zusätzlich muss auch die Wahlbeteiligung der 17. Bundestagswahl nochmal regressiert werden. Für die

¹⁰ Aus Gründen des Platzmangels wurden sämtliche Tabellen, die das Verfahren beschreiben, in der Datei „Regressoren.docx“ hinterlegt.

Regression Erststimme FDP (*FDP1*) müssen die zwei Regressionskoeffizienten *age1* und *edu2* entnommen werden, da beide mit anderen Variablen aus denselben Gruppen korrelieren und sie die niedrigste Toleranz beziehungsweise den höchsten VIF-Wert aufweisen. Die gleiche Vorgehensweise wird auch für die anderen Wahljahre angewendet. SPSS hat mit der Rückwärtsregression für das Jahr 2002 selten die Prädiktoren *sex*, *mig*, *une*, *age1*, *age3* und *age4* mittels *F*-Wahrscheinlichkeitsuntersuchung ausgeschlossen. Öfter eliminiert sind *age2*, *age5*, *se1* und *sec4*. Daten von *sec2* und *sec3* werden sehr häufig aus den Modellen entnommen. Die Variablengruppen, die die Bildung wiedergeben (*edu1*, *edu2*, *edu3* und *edu4*) werden nicht von SPSS ausgeschlossen. Wegen Kollinearität muss in den Modellen stets eine dieser aussortiert werden. Dabei weist der Regressionskoeffizient *edu2* stets auf Multikollinearität. Dies gilt auch für das Jahr 2005 ausschließlich der Regression der Wahlbeteiligung. *Edu1*, *edu3* und *edu4* werden von SPSS gar nicht bis maximal einmalig ausgeschlossen. Die Prädiktoren *sex* wurde ebenfalls nie entnommen. Selten entfernt werden *mig*, *une*, *age2*, *age3*, *age4* und *age5*; auf *age1* wird vier Mal verzichtet. Die Gruppen der Sektoren weisen durchmischte Ergebnisse auf. Auffällig ist, dass auch in diesem Jahr wiederholt mehrfach *sec2* und *sec3* isoliert werden. Für die Regressionsmodelle der 17. Bundestagswahl (2009) werden eher die Regressionskoeffizienten *age2* und *age4* eliminiert. Für neu zu regressierende Modellen muss zusätzlich die Variable *edu4* entfernt werden. Unerwarteter Weise haben die Regressionsmodelle für Erst- und Zweitstimme der Die Linken trotz Einbeziehung aller Bildungsgruppen *edu1*, *edu2*, *edu4* und *edu4* sehr gute Toleranz- und VIF-Werte. Obwohl dies der Theorie widerspricht werden diese dennoch beibehalten um das Modell möglichst nicht zu verfälschen. 2013 müssen nur Erst- und Zweitstimmen der Die Linken wiederholt regressiert werden. SPSS hat für beide Modelle die Prädiktoren *edu1*, *sec1*, *sec2* ausgeschlossen. *Age5* muss zusätzlich auch aus dem Modell entnommen werden.

Da bei allen Regressionen die Rückwärts-Methode angewendet wird, existieren ausgeschlossene Variablen, deren Kennwerte nun beschrieben werden. Diese Prädiktoren werden mit dem *t*-Test getestet, um die Signifikanz der Beziehung zur unabhängigen Variablen zu entdecken. Die einzelnen Tests für jede unabhängige Variable zeigen, dass all jene ein zu hohes Signifikanzniveau besitzen. Aus diesem Grund werden die aufgezeichneten Variablen zu Recht ausgewiesen. Der Durchschnitt aller Signifikanzniveaus liegt bei 0,676 und verweist auf hohe Alpha-Fehler beim *t*-Test. Eine Beziehung dieser unabhängigen Variablen zur abhängigen Variablen kann also nicht garantiert werden. Beim Vergleich der

Anzahl der ausgeschlossenen Variablen ist auffällig, dass im Jahr 2013 die meisten Prädiktoren eliminiert sind. Darauf folgend verstößt SPSS in den Jahren 2002 und 2005 die nächst größere Menge an unabhängigen Variablen. Die geringste Variableneliminierung passiert in Regressionen des 17. Bundestagswahl 2009. Hier sollte jedoch beachtet werden, dass ohnehin die Daten zu den sozialpflichtigen in den unterschiedlichen Sektoren (*sec1*, *sec2*, *sec3* und *sec4*) nicht gegeben sind. Diese gelten also generell als ausgeschlossen.

Die erneute Regression wird nun mit der Einschluss-Methode durchgeführt, um Multikollinearität zu beseitigen. Anhand der Toleranz und VIF-Werte der neuern Modelle wird beobachtet, dass Multikollinearität durch die Eliminierung der betroffenen Regressionskoeffizienten deutlich zurückgeht. Die Beobachtungen der Regressionen bezüglich den globalen Prüfungen der Modelle und den Prüfungen der einzelnen Regressionskoeffizienten werden im nächsten Abschnitt beschrieben und dann ausgewertet.

3.3 Modellprüfung

Zuerst werden die ANOVAs analysiert. Im Anschluss daran werden Kennzahlen zur Bewertung der Modelle in den einzelnen Bundestagswahlen untersucht.

Zur Modellprüfung wird auch ANOVA (analysis of variance) genutzt – eine Menge an statistischen Daten, die zusätzlich Informationen zum Gesamtmodell geben (vgl. Urban, 2006).¹¹ Da der F -Wert zuvor analysiert wurde, wird dieser nicht mehr im ANOVA betrachtet. Die Reihe „Regression“ gibt die erklärte Streuung ($\sum_{k=1}^K (\hat{y}_k - \bar{y})^2$) und Reihe „Residuen“ die ungeklärte Streuung ($\sum_{k=1}^K (y_k - \hat{y}_k)^2$) wider. „Gesamtsumme“ beinhaltet die Gesamtvarianz ($\sum_{k=1}^K (y_k - \bar{y})^2$) (vgl. Urban, 2006). Von 44 Regressionen haben 26 Modelle ein Mittel der quadrierten Streuung Ebenfalls in der Tabelle zu finden sind die F -Werte der Modelle.

Ein Vergleich der Kennwerte zwischen den mit Multikollinearität behafteten Modellen und den davon gesäuberten Modellen zeigt, dass die älteren Regressionsmodelle stets ein höheres Bestimmtheitsmaß besitzen. Die Differenzen der R-Quadrate sind meist gering. Durchschnittlich beträgt diese 0,042 was 4,2 Prozent entspricht. Dies bedeutet, dass die neuen Modelle durchschnittlich 4,2% der Varianz in y durch die Varianz in x weniger klärt

¹¹ Ergebnisse bzw. Tabellen werden in der Datei „ANOVA.docx“ präsentiert

werden kann als mit den vorherigen Regressionsmodellen. Es kann auch beobachtet werden, dass das Bestimmtheitsmaß bei jeder Regression größer ist als das adjustierte Bestimmtheitsmaß. Wie es dazu kommt wurde bereits im Theorieteil erklärt. Die folgende Tabelle 3.2 verschafft eine Übersicht über globale Kennzeichen aller ausgewählten Modelle.

Die Regressionsergebnisse der 15. Bundestagswahl lassen ein hohes Bestimmtheitsmaß aufweisen. Nur das Regressionsmodell für den Erststimmenanteil der FDP (FDP1) besitzt ein Bestimmtheitsmaß der weniger als 30 Prozent beträgt. Die in Farbe Grau markierten Felder zeigen diejenigen R^2 und \bar{R}^2 , die im Vergleich zwischen den Erst- und Zweitstimmenanteil einen höheren Wert aufweist. Mit Ausnahme der SPD, weisen bei allen anderen Parteien im Jahr 2002 die Zweitstimme einen höheren Bestimmtheitsmaß auf. Dies ist auch gültig für das adjustierte Bestimmtheitsmaß. Zur Bewertung des Fischer-Tests müssen Tabellen der F -Statistik angeschaut werden. Alle von SPSS gegebenen empirischen F -Werte der Modelle für Wahljahr 2002 sind größer als die theoretischen F -Werte. In den Fischer-Tests mit 95 prozentiger Sicherheit werden somit die Nullhypothesen $H_0: \beta_1 = \beta_2 = \dots = \beta_k = \dots = \beta_K = 0$ abgelehnt. Folglich wird die Vermutung, dass alle einbezogenen unabhängigen Variablen keinen Einfluss auf die abhängige Variable haben, verworfen. Das Signifikanzniveau des Fischer-Tests ist dabei stets rund Null – also ist dieser höchst signifikant. Die Standardfehler der Schätzmodelle befinden sich in Spalte „ σ Schätzung“ und geben wie bereits erwähnt die Abweichungen zwischen den beobachteten und den geschätzten y -Variablen. Im Jahr 2002 ist der Standardfehler für Zweitstimmen – außer für CDU/CSU – kleiner als der der Erststimmen.

Die Regressionen der Wahlbeteiligung und der Wahlergebnisse des Wahljahrs 2005 liefern ähnlich gute Bestimmtheitsmaße und adjustierte Bestimmtheitsmaße. Das Modell für Erststimme FDP liefert wiederholt schlechter als 30 Prozent Güte. Bis auf SPD und CDU haben andere Parteien eine mit der Regression der Zweitstimme ein höheres \bar{R}^2 . Die Fischer-Tests sind ebenfalls hier hoch signifikant, sodass H_0 abgelehnt werden kann. Zum Teil sind die F -Werte so ausgeprägt, dass die Nullhypothese binnen kürzester Beobachtungszeit verworfen werden kann. Ausgenommen der Stimmenanteile der Partei SPD sind die Standardfehler der anderen Parteien bei Regression der Zweitstimme höher als Erststimme.

Modellübersicht^b

	Modell	R- Quadrat	Angepasstes R-Quadrat	S Schätzung	F-Wert	Sig. Änderung in F	Durbin- Watson
2002	WB	,755	,746	1,952264	88,752	,000	1,310
	CDU/CSU1	,613	,604	7,094857	65,874	,000	,795
	CDU/CSU2	,620	,610	7,169822	59,268	,000	,669
	SPD1	,383	,359	7,431024	16,172	,000	,817
	SPD2	,366	,344	6,834860	16,658	,000	,622
	FDP1	,181	,155	1,550957	7,075	,000	1,483
	FDP2	,495	,477	1,397240	28,203	,000	,801
	LIN1	,823	,816	3,244469	121,008	,000	1,364
	LIN2	,865	,861	2,397343	185,219	,000	1,411
	GRÜ1	,525	,507	2,163973	28,816	,000	1,612
	GRÜ2	,784	,777	1,966238	116,434	,000	1,005
2005	WB	,585	,569	1,942482	36,738	,000	1,452
	CDU/CSU1	,634	,620	6,522703	45,249	,000	,925
	CDU/CSU2	,633	,617	5,815911	41,080	,000	,816
	SPD1	,365	,340	7,747758	14,980	,000	,871
	SPD2	,331	,305	6,423434	12,905	,000	,791
	FDP1	,174	,142	1,337842	5,500	,000	1,493
	FDP2	,456	,439	1,688097	26,890	,000	,957
	LIN1	,861	,855	3,410844	147,905	,000	1,352
	LIN2	,896	,892	2,767958	225,361	,000	1,181
	GRÜ1	,507	,490	2,512184	29,647	,000	1,550
	GRÜ2	,768	,759	1,789600	86,164	,000	,935
2009	WB	,575	,562	2,848842	43,414	,000	1,397
	CDU/CSU1	,580	,569	5,466868	50,124	,000	1,255
	CDU/CSU2	,583	,573	4,412324	58,110	,000	,994
	SPD1	,414	,393	6,736441	20,313	,000	1,030
	SPD2	,437	,418	4,925073	22,363	,000	,864
	FDP1	,302	,288	2,023402	21,097	,000	1,429
	FDP2	,524	,508	2,330760	31,745	,000	,849
	LIN1	,817	,811	3,977599	142,969	,000	1,385
	LIN2	,858	,854	3,233536	194,125	,000	1,283
	GRÜ1	,439	,426	3,479900	32,538	,000	1,691
	GRÜ2	,657	,647	2,452591	61,556	,000	1,178
2013	WB	,253	,235	3,393270	14,065	,000	1,434
	CDU/CSU1	,156	,127	8,183259	5,341	,000	,998
	CDU/CSU2	,177	,148	6,841380	6,177	,000	,920
	SPD1	,515	,494	6,571531	25,260	,000	1,168
	SPD2	,550	,532	5,030508	29,180	,000	1,066
	FDP1	,076	,063	,892977	6,034	,000	1,632
	FDP2	,317	,300	1,261595	19,286	,000	,906
	LIN1	,404	,379	6,361153	16,168	,000	,548
	LIN2	,405	,380	5,534298	16,191	,000	,442
	GRÜ1	,116	,104	3,685576	9,630	,000	1,317
	GRÜ2	,178	,164	3,078724	12,719	,000	1,085

Tabelle 3.2: Güte der auserwählten Modelle für die Bundestagswahlen 2002, 2005, 2009 und 2013.

Die 17. Bundestagswahl gibt solide Ergebnisse bezüglich die Güte der Modelle. Sämtliche R^2 -Werte liegen über 0,3. Wie in den vorherigen Jahren hat Die Linke bezüglich dessen die besten Werte der adjustierten Bestimmtheitsmaße. F -Werte fallen erneut bemerkbar hoch und signifikant aus. Es kann wieder ausgeschlossen werden, dass sämtliche in die Regression eingehenden Variablen keinen Einfluss haben. Bis auf FDP sind die σ aller anderen Parteien Parteien bei der Zweitstimmen- geringer als die der Erststimmenanteile. In den Modellen der 18. Bundestagswahl sind für die Parteien SPD und Die Linke gute adjustierte Bestimmtheitsmaße zu finden. Restliche Wahlergebnisse anderer Parteien und Wahlbeteiligung weisen auf eine sehr geringe Güte. Insgesamt fallen in diesem Wahljahr die F -Werte geringer aus. Dennoch können damit H_0 verworfen werden. Signifikanzfehler treten durchwachsen und unterschiedlich stark aus.

Regressionen für die 15., 16. Und 17. Bundestagswahlen liefern überwiegend eine hohe Güte. Zum 18. Bundestagswahl jedoch, sinken die \bar{R}^2 -Werte deutlich (siehe Tabelle 3.2), obwohl dieselben unabhängigen Variablen und dieselbe Grundgesamtheit zur Analyse gezogen werden. Im Mittel liegt R^2 lediglich bei 0,286. Nur Regressionen der Stimmenanteile der SPD haben ein höheres adjustiertes Bestimmtheitsmaß. Besonders auffällig ist die Regression der Erststimme der Partei FDP. Über die Jahre blieb dessen adjustiertes Bestimmtheitsmaß von allen Modellen das niedrigste. Durchweg solide Werte liefern Regressionen der Erst- und Zweitstimmen der CDU/CSU, der SPD und der Grünen.

3.4 Prüfung der Regressionskoeffizienten

Die von den Modellen ausgeschlossenen Variablen wurden bereits im Kapitel 3.2 beschrieben. Nun werden die Regressionskoeffizienten so genau wie möglich untersucht und charakterisiert. Zur Prüfung der Koeffizienten werden t -Tests, Standardfehler und Konfidenzintervalle genutzt.

Bei den Regressionskoeffizienten wird die Nullhypothese im t -Test verworfen, wenn der empirische t -Wert stark von Null abweicht. Das Verwerfen von H_0 wird als Existenz eines Zusammenhangs zwischen den x - und y -Variablen interpretiert. Regressionskoeffizienten, die einen t -Wert um Null besitzen, sind orange markiert. Relevante Regressionskoeffizienten mit einem höheren Signifikanzwert (größer als 0,05) bezüglich t -Test sind in im Anhang von

Tabelle 6.3 bis Tabelle 6.19 orange gekennzeichnet. Koeffizienten, die über einen zu niedrigen *t*-Wert verfügen, haben auch einen schlechteren Sig.-Wert.

Zunächst werden die Regressionskoeffizienten auffällig guter oder schlechter Modelle im Jahr 2002 analysiert und interpretiert. Die Wahlbeteiligung einem Wahlkreis sinkt um 0,08 Prozent, wenn der Anteil der männlichen Bevölkerung in diesem Wahlkreis um ein Prozent steigt. Zugleich senkt eine ein-prozentige Steigung der nicht-deutschen Bevölkerung die Wahlbeteiligung um 0,17 Prozent. Wächst die Arbeitslosenquote um ein Prozent, wird die Wahlbeteiligung um 0,42 Prozent negativ beeinträchtigt – ähnlich auch der Einfluss des Regressors *sec1*, bei dem die Wahlbeteiligung um -0,47 verändert wird. Geringe positive Einflüsse auf die Wahlbeteiligung werden nach dem Modell durch Anteil der unter 18-Jährigen und dem Anteil der Schulabsolventen mit allgemeiner Hochschulreife gegeben. Die größte negative Beeinflussung verursacht *age2*, dem Anteil der 18 bis 25 Jährigen Bevölkerung in einem Wahlkreis. Durch diese sinkt die Wahlbeteiligung um 1,54 Prozent. Der Anteil der Altersgruppe der 35 bis 60-Jährigen scheint einen positiven Einfluss von 0,66 Prozent zu haben. All die eben genannten Koeffizienten sind signifikant bei einem 10 und 5-prozentigen Niveau. Der Koeffizient *edu3* ist weder bei 1, 5 noch 10-prozentigem Niveau signifikant.

Die Erststimmenanteile der CDU/CSU werden von allen einbehaltenen Koeffizienten nach der Rückwärtsregression negativ beeinflusst. Ein relativ hoher und signifikanter Einfluss von 0,89 Prozent trägt die Arbeitslosenquote bei. Ähnlich ist es auch bei der Zweitstimme. Werden die standardisierten verglichen, kann festgestellt werden, dass Arbeitslosigkeit am stärksten negativen Einfluss auf Stimmenanteile der CDU/CSU nimmt. Diese werden aus der Tabelle 6.3 entnommen.

Bei der SPD wirken die meisten Regressoren dagegen positiv auf die Stimmenanteile. Ein besonders positiver Einfluss haben die Variablen *age5*, *age1*, *age3*, *age4* und *edu4*. Steigt der Anteil der über 60-jährigen Bevölkerung um 1 Prozent, vergrößert sich der Anteil der SPD-Stimmen in einem Wahlkreis um 4,86 Prozent. Der Koeffizient *sex* – also Anteil der männlichen Bevölkerung hat bezüglich SPD-Stimmenanteile leicht negative aber nicht-signifikante Einflüsse auf Erst- und Zweitstimmen.

Die geringen Einflüsse der Regressoren auf die Zweitstimmenanteile der FDP haben fast alle Sig.-Werte, die kleiner als 0,05 sind. Nur *sec4* ist bei einem 10 prozentigem Niveau nicht

signifikant. Dies wird in der Tabelle 6.5 angezeigt. Das Modell der Erststimmenanteile der FDP besitzt hingegen fünf von insgesamt neun nicht-signifikante Koeffizienten. Die Variablen haben keinen signifikanten Einfluss auf die Erststimmenanteile der FDP.

Bei der Linken sind Altersgruppen starke negative Einflussfaktoren auf Erst- und Zweitstimmenanteile. Im Vergleich zu ihren Einflüssen auf Stimmen der SPD, sind die der Linken geringer. Die Werte der standardisierten Regressionskoeffizienten liegen bei der Sozialdemokratischen Partei Deutschlands deutlich höher als bei der Linken.

Dieselben Koeffizienten bezüglich Anteile der Schulabsolventen der Zweitstimmenanteile der Grünen sind weniger signifikant als bei den Erststimmenanteilen. Regressoren *mig*, *age1*, *age3*, *age4* und *age5*, die in beiden Fällen signifikant Einfluss nehmen, haben bei der Erststimme einen leicht größeren und vor allem negativen Einfluss.

Auch im Jahr 2005 kommt der größte Einfluss auf die Wahlbeteiligung vom Regressionskoeffizienten *age2*, der dazu sehr hoch signifikant ist. Die Stärke und Richtung der Einflüsse dieser Variablen spiegelt sich in den geringen standardisierten Regressionskoeffizienten wider. Alle anderen unabhängigen signifikanten Variablen beeinträchtigen die Wahlbeteiligung weniger als um 0,4 Prozent. Regressionskoeffizienten der Stimmenanteile einzelner Parteien zum 16. Bundestagswahl befinden sich in Tabelle 6.7.

Bei allen Parteien lassen sich keine signifikante Beziehung zwischen dem Anteil der Schulabgänger ohne (Haupt-) Schulabschluss (*edu1*) vermuten. Verglichen mit den standardisierten Koeffizienten im Jahr 2002 haben die Regressoren dieses Wahljahr einen geringeren Wert. Die Signifikanzniveaus der Regressoren sind auch weniger signifikant und erschweren damit die Interpretation der Einflüsse der x-Variablen.

Bei den Regressionsmodellen im Wahljahr 2009 können die Gruppen der Sozialversicherungspflichtigen in den Sektoren nicht einbezogen werden, da diese Daten nicht zur Verfügung stehen. Variable *age2* bleibt weiterhin der Regressor mit dem größten Einfluss auf die Wahlbeteiligung. Wie die letzten Jahre ist der Wert auch diesmal hoch signifikant. Dies lässt sich in Tabelle 6.12 zeigen.

Signifikanzniveaus der Regressionskoeffizienten in kleineren Parteien fallen erneut gering aus. Die Variablen mit einer schlechten Signifikanz sind in den Tabelle 6.13 bis Tabelle 6.15

orange markiert. Währenddessen ist die Signifikanz der Großparteien CDU/CSU und SPD durchgehend hoch, wie man anhand der Tabelle 6.12 und Tabelle 6.13 erkennen kann.

Die vielen orange markierten Koeffizienten in den Tabelle 6.16 bis Tabelle 6.19 weisen auf nicht genügende Signifikanz der Regressionskoeffizienten hin. Beim Vergleich der standardisierten Variablen fällt auf, dass diese deutlich geringer sind als die der letzten Modelle. Dies zeigt, dass die Koeffizienten meist einen kleinen Einfluss auf die abhängigen Variablen haben.

4 Schluss

Bei dieser Arbeit wurde der Zusammenhang zwischen Strukturdaten und Stimmenanteile einzelner Parteien anhand einer multivariaten linearen Regression untersucht. Dabei wurden ausgewählte Strukturdaten betrachtet, wo sich ein linearer Zusammenhang zu den Wahlergebnissen vermuten lässt. Variablen, die Geschlecht (*sex*), Alter (*age1* bis *age5*), Bildung (*edu1* bis *edu4*), Arbeitslosigkeit (*une*), Einwanderung (*mig*) und Beschäftigtenanteile in Wirtschaftssektoren (*sec1* bis *sec4*) beschreiben einen Wahlkreis.

Die erstellten linearen Modelle und deren Güte konnten anhand von Kennzahlen bewertet werden. Indem unterschiedlich relevante Regressoren eingebracht werden, zeigt das multiple lineare Modell bei jedem Wahljahr eine unterschiedliche Güte. Das Modell konnte die Wahlergebnisse in 2002 für alle Parteien sehr gut erklären, sodass ein linearer Zusammenhang zwischen den abhängigen Variablen und die Anteile der Stimmen erklärt werden kann. Für 2005 erhalten die Regressionen der Stimmenanteile solide Modellgüte. Die Werte blieben weiterhin für die Regression der 17. Bundestagswahl gut erhalten, wobei insbesondere die Erst- und Zweitstimmenanteile der Linken gut erklärt werden kann. Regressoren, die die Altersstruktur beschreiben, nehmen starken negativen Einfluss auf das Wahlergebnis der Linken. In diesen drei Bundestagswahlen konnte die Regression der Wahlbeteiligung eine Güte von über 0,5 erzielen. Die größten Einflüsse auf die Wahlbeteiligung werden durch Altersgruppen und Bildungsgruppen ausgeübt. Die Modellgüte der anderen Parteien können ebenfalls als gut bewertet werden. Die Ergebnisse der 18. Bundestagswahl konnten jedoch deutlich weniger gut durch einen linearen Zusammenhang erklärt werden, obwohl dieselben Daten als erklärenden Variablen in die Untersuchung einfließen. 2013 zeigt sich ein starker Einriss der Modellgüte.

Es kann ebenfalls beobachtet werden, dass das multiple lineare Regressionsmodell Anteile der Zweitstimmen besser erklären kann als die der Erststimme. Dies gilt sowohl für kleinere als auch größere Parteien wie CDU/CSU oder SPD. Mit der Erststimme wird der Politiker (einer Partei) gewählt. Es zeigt sich, dass eher ein Zusammenhang zwischen den Stimmen der Parteien und den Strukturdaten besteht als zwischen Stimmen für einen bestimmten Kandidaten und den Strukturdaten.

Wie in der Einleitung angedeutet, konnte durch die Untersuchung ein linearer Zusammenhang zwischen den Strukturdaten und den Wahlergebnissen eines Wahlkreises nachgewiesen werden. Um noch genauere Regressionsmodelle zu erhalten, sind zusätzliche Informationen nützlich. Daten über Einkommen, Bildungsniveaus der Gesamtbevölkerung in einem Wahlkreis (nicht wie untersuchten Fall: Schulabgängern aus vorherigen Jahren) und weitere soziale, geographische und wirtschaftliche Informationen schaffen ein noch besseres Regressionsmodell. Mit diesem Ansatz können anhand von sozio-ökonomischen Bedingungen Wahlbeteiligung und Wahlergebnisse vorhergesagt werden.

5 Literatur

- AGCI, S. (2010): „Wählerschaft und Mitglieder“ in „Parteien in Deutschland“, <http://www.bpb.de/politik/grundfragen/parteien-in-deutschland/42159/waehlerschaft-und-mitglieder> (14.12.14, 18:25).
- ARBEITSAGENTUR (2014): „Arbeitslosenquoten“ <https://statistik.arbeitsagentur.de/Navigation/Statistik/Grundlagen/Berechnung-der-Arbeitslosenquote/Berechnung-der-Arbeitslosenquote-Nav.html> (25.12.14, 21:03).
- AUER, V.L. (2011): „Ökonometrie“, 5, Heidelberg.
- BACKHAUS, K./E.,B./P.,W./R.,R. (2011): „Multivariate Analysemethoden“, „Eine anwendungsorientierte Einführung“, Berlin, 5, S. 55 – 117.
- BECKMANN, M. (2014): „Lässt sich der Ausgang von politischen Wahlen durch die Analyse von Twitterdaten vorhersagen?“, Berlin.
- BUNDESWAHLLEITER (2014): „Strukturdaten für die Wahlkreise zum 17. Deutschen Bundestag“ http://www.bundeswahlleiter.de/de/bundestagswahlen/BTW_BUND_09/strukturdaten/ (25.12.2014, 20:19 Uhr).
- BUNDESWAHLLEITER (2013): „Bundestagswahl 2013“, „Durchführung der Repräsentativen Wahlstatistik“ http://www.bundeswahlleiter.de/de/bundestagswahlen/BTW_BUND_13/veroeffentlichungen/BTW13_FB_ReprWStat.pdf (24.12.14, 18:40).
- BUNDESWAHLLEITER, (2010): Wahl zum 17. Deutschen Bundestag am 27. September 2009, Heft 4: Wahlbeteiligung und Stimmabgabe der Männer und Frauen nach Altersgruppen, Wiesbaden.
- BUNDESWAHLLEITER (2013): „Allgemeine und methodische Hinweise zur repräsentativen Bundeswahlstatistik 2013“ http://www.bundeswahlleiter.de/de/bundestagswahlen/BTW_BUND_13/veroeffentlichungen/repraesentative/methodische_hinweise.pdf (24.12.14, 18:40).
- CDU (2014) : „Grundsatzprogramm der CDU Deutschlands“, „Freiheit und Sicherheit. Grundsätze für Deutschland.“ <http://www.cdu.de/grundsatzprogramm> (12.12.2014, 22:50 Uhr). „Freiheit und Sicherheit“, „Grundsätze für Deutschland“ http://www.cdu.de/sites/default/files/media/dokumente/071203-beschluss-grundsatzprogramm-6-navigierbar_1.pdf (14.12.2014, 17:00 Uhr).
- Freiheit und Sicherheit. Grundsätze für Deutschland. Das Grundsatzprogramm, Berlin.
- DEUTSCHER BUNDESTAG (2014): „III. Der Bundestag“ http://www.bundestag.de/bundestag/aufgaben/rechtsgrundlagen/grundgesetz/gg_03/ (14.12.2014, 17:00 Uhr).
- FAHRMEIR, L./ H., A./ T., G. (1996): „Multivariate statistische Verfahren“, 2, Berlin

- GRUNDGESETZ (2013): Artikel 38 Absatz 1, Artikel 39 Absatz 1.
- HARTUNG, J./E.B. (1999): „Multivariate Statistik“, „Lehr. Und Handbuch der Angewandten Statistik“, 6, München.
- KÄHLER, W.-M. (1993): „Statistische Datenanalyse mit SPSS/PC+“, 3, Wiesbaden.
- KELLER, D. (2013): „Analyse von Zusammenhängen: Korrelation“ <http://www.statistik-und-beratung.de/2013/05/analyse-von-zusammenhangen-korrelation/> (24.12.14, 18:15)
- KOCH, S. (1994): „Parteien in der Region“, „Mitglieder Wahlergebnisse Parteienstruktur“.5, München.
- KORTMANN, M. (2010): „Wählerschaft und Mitglieder“ in „Parteien in Deutschland“, <http://www.bpb.de/politik/grundfragen/parteien-in-deutschland/42117/waehlerschaft-und-mitglieder>(14.12.14, 18:15)
- KRAUSE, A. (1994): „Computerintensive statistische Methoden“, Stuttgart.
- LESER, U. (2014): „Lässt sich der Ausgang von politischen Wahlen durch die Analyse von Twitterdaten vorhersagen?“, Berlin.
- MERTENS, M. (2010): „Wählerschaft und Mitglieder“ in „Parteien in Deutschland“, <http://www.bpb.de/politik/grundfragen/parteien-in-deutschland/42093/waehlerschaft-und-mitglieder> (14.12.14, 18:00)
- NIEDERMAYER, O. (2012): „Nationaldemokratische Partei Deutschlands“ in „Parteien in Deutschland“, <http://www.bpb.de/politik/grundfragen/parteien-in-deutschland/42205/nationaldemokratische-partei-deutschlands> (14.12.14, 18:30)
- Parteimitglieder in Deutschland: Version 2008. Arbeitshefte a. d. Otto-Stammer-Zentrum, Nr. 13, FU Berlin 2008.
- OLZOG, G. (1999): „Die politischen Parteien in Deutschland“, „Geschichte – Programmatik – Organisation – Personen – Finanzierung“. 25, München.
- ROSAR, U. (2009): „Fabulous Front-Runners. Eine empirische Untersuchung zur Bedeutung der physischen Attraktivität von Spitzenkandidaten für den Wahlerfolg ihrer Parteien“ in PVS Politische Vierteljahrszeitschrift, S. 1 – 4.
- SCHNEIKER, A. (2010): „Wählerschaft und Mitglieder“ in „Parteien in Deutschland“, Bundeszentrale für politische Bildung. <http://www.bpb.de/politik/grundfragen/parteien-in-deutschland/42068/waehlerschaft-und-mitglieder> (14.12.14, 17:30)
- SONTHHEIMER, K./B., W./G., Andrea (2007): „Grundzüge des politischen Systems Deutschlands“, München.
- STATISTA (2014): „Statistik-Lexikon: Definition Regressionsanalyse“, <http://de.statista.com/statistik/lexikon/definition/113/regressionsanalyse/> (13.12.14, 21:00)

TAKEN, J. (2010): „Wählerschaft und Mitglieder“ in „Parteien in Deutschland“, (14.12.14, 18:20)

URBAN, D. (1982): „Regressionstheorie und Regressionstechnik“, Duisburg.

URBAN, D./M.,J. (2006): „Regressionsanalyse: Theorie, Technik und Anwendung“, 2, Wiesbaden.

WALTER, Franz. (2004): Zurück zum alten Bürgertum: CDU/CSU und FDP, in: Aus Politik und Zeitgeschichte, Heft 40/2004, 32-38.

WELLER, K.(2014): „Twitter and society“, New York.

WOOLDRIDGE, J. M. (2013); „Introductory Econometrics“, „A Modern Aproach“, (S. 67), 5. Aufl., Kanada.

Sämtliche Daten aus [www. Bundeswahlleiter.de](http://www.Bundeswahlleiter.de)

6 Anhang

Tabelle 6.1: Test auf Multikollinearität mit Spearman-Korrelationskoeffizienten.

		Korrelationen															
		sex	mig	une	age1	age2	age3	age4	age5	edu1	edu2	edu3	edu4	sec1	sec2	sec3	sec4
sex	Pearson-Korrelation	1	,067*	,022	,017	,007	,022	,093**	-,092**	,019	,018	,014	-,028	,032	-,006	,035	-,012
	Sig. (2-seitig)		,021	,440	,547	,796	,451	,001	,001	,522	,536	,624	,326	,342	,857	,295	,722
	N	1196	1196	1196	1196	1196	1196	1196	1196	1196	1196	1196	1196	897	897	897	897
mig	Pearson-Korrelation	,067*	1	-,246**	-,053	-,069*	,534**	-,140**	-,294**	-,180**	,151**	-,481**	,237**	-,596**	-,224**	,038	,306**
	Sig. (2-seitig)	,021		,000	,069	,018	,000	,000	,000	,000	,000	,000	,000	,000	,000	,251	,000
	N	1196	1196	1196	1196	1196	1196	1196	1196	1196	1196	1196	1196	897	897	897	897
une	Pearson-Korrelation	,022	-,246**	1	-,074*	,052	-,453**	,442**	,169**	-,070*	,008	,241**	-,173**	,500**	-,328**	-,006	,297**
	Sig. (2-seitig)	,440	,000		,010	,072	,000	,000	,000	,016	,774	,000	,000	,000	,000	,857	,000
	N	1196	1196	1196	1196	1196	1196	1196	1196	1196	1196	1196	1196	897	897	897	897
age1	Pearson-Korrelation	,017	-,053	-,074*	1	-,190**	,018	-,444**	-,617**	-,134**	,614**	,094**	-,623**	,021	,574**	-,094**	-,591**
	Sig. (2-seitig)	,547	,069	,010		,000	,532	,000	,000	,000	,000	,001	,000	,534	,000	,005	,000
	N	1196	1196	1196	1196	1196	1196	1196	1196	1196	1196	1196	1196	897	897	897	897
age2	Pearson-Korrelation	,007	-,069*	,052	-,190**	1	,125**	-,032	-,242**	,013	-,219**	,014	,216**	,054	-,193**	-,165**	,272**
	Sig. (2-seitig)	,796	,018	,072	,000		,000	,268	,000	,661	,000	,629	,000	,109	,000	,000	,000
	N	1196	1196	1196	1196	1196	1196	1196	1196	1196	1196	1196	1196	897	897	897	897
age3	Pearson-Korrelation	,022	,534**	-,453**	,018	,125**	1	-,515**	-,592**	,230**	,132**	-,454**	,124**	-,348**	-,266**	-,127**	,388**
	Sig. (2-seitig)	,451	,000	,000	,532	,000		,000	,000	,000	,000	,000	,000	,000	,000	,000	,000
	N	1196	1196	1196	1196	1196	1196	1196	1196	1196	1196	1196	1196	897	897	897	897
age4	Pearson-Korrelation	,093**	-,140**	,442**	-,444**	-,032	-,515**	1	,258**	-,079**	-,303**	,105**	,251**	,247**	-,260**	,267**	,143**
	Sig. (2-seitig)	,001	,000	,000	,000	,268	,000		,000	,006	,000	,000	,000	,000	,000	,000	,000
	N	1196	1196	1196	1196	1196	1196	1196	1196	1196	1196	1196	1196	897	897	897	897
age5	Pearson-Korrelation	-,092**	-,294**	,169**	-,617**	-,242**	-,592**	,258**	1	-,024	-,428**	,227**	,253**	,107**	-,091**	,093**	,047
	Sig. (2-seitig)	,001	,000	,000	,000	,000	,000			,401	,000	,000	,000	,001	,006	,005	,156
	N	1196	1196	1196	1196	1196	1196	1196	1196	1196	1196	1196	1196	897	897	897	897

Tabelle 6.2: Test auf Multikollinearität mit Spearman-Korrelationskoeffizienten.

		Korrelationen															
		sex	mig	une	age1	age2	age3	age4	age5	edu1	edu2	edu3	edu4	sec1	sec2	sec3	sec4
edu1	Pearson-Korrelation	,019	-,180**	-,070	-,134**	,013	,230**	-,079**	-,024	1	-,009	,008	-,295**	,370**	-,147**	-,042	,130**
	Sig. (2-seitig)	,522	,000	,016	,000	,661	,000	,006	,401		,745	,793	,000	,000	,000	,209	,000
	N	1196	1196	1196	1196	1196	1196	1196	1196	1196	1196	1196	1196	897	897	897	897
edu2	Pearson-Korrelation	,018	,151**	,008	,614**	-,219**	,132**	-,303**	-,428**	-,009	1	-,365**	-,653**	-,212**	,467**	-,104**	-,440**
	Sig. (2-seitig)	,536	,000	,774	,000	,000	,000	,000	,000	,745		,000	,000	,000	,000	,002	,000
	N	1196	1196	1196	1196	1196	1196	1196	1196	1196	1196	1196	1196	897	897	897	897
edu3	Pearson-Korrelation	,014	-,481**	,241**	,094**	,014	-,454**	,105**	,227**	,008	-,365**	1	-,393**	,514**	,270**	-,072**	-,332**
	Sig. (2-seitig)	,624	,000	,000	,001	,629	,000	,000	,000	,793	,000		,000	,000	,000	,031	,000
	N	1196	1196	1196	1196	1196	1196	1196	1196	1196	1196	1196	1196	897	897	897	897
edu4	Pearson-Korrelation	-,028	,237**	-,173**	-,623**	,216**	,124**	,251**	,253**	-,295**	-,653**	-,393**	1	-,275**	-,562**	,136**	,594**
	Sig. (2-seitig)	,326	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000		,000	,000	,000	,000
	N	1196	1196	1196	1196	1196	1196	1196	1196	1196	1196	1196	1196	897	897	897	897
sec1	Pearson-Korrelation	,032	-,596**	,500**	,021	,054	-,348**	,247**	,107**	,370**	-,212**	,514**	-,275**	1	,044	,044	-,194**
	Sig. (2-seitig)	,342	,000	,000	,534	,109	,000	,000	,001	,000	,000	,000	,000		,189	,190	,000
	N	897	897	897	897	897	897	897	897	897	897	897	897	897	897	897	897
sec2	Pearson-Korrelation	-,006	-,224**	-,328**	,574**	-,193**	-,266**	-,260**	-,091**	-,147**	,467**	,270**	-,562**	,044	1	-,443**	-,916**
	Sig. (2-seitig)	,857	,000	,000	,000	,000	,000	,000	,006	,000	,000	,000	,000	,189		,000	,000
	N	897	897	897	897	897	897	897	897	897	897	897	897	897	897	897	897
sec3	Pearson-Korrelation	,035	,038	-,006	-,094**	-,165**	-,127**	,267**	,093**	-,042	-,104**	-,072**	,136**	,044	-,443**	1	,066*
	Sig. (2-seitig)	,295	,251	,857	,005	,000	,000	,000	,005	,209	,002	,031	,000	,190	,000		,048
	N	897	897	897	897	897	897	897	897	897	897	897	897	897	897	897	897
sec4	Pearson-Korrelation	-,012	,306**	,297**	-,591**	,272**	,388**	,143**	,047	,130**	-,440**	-,332**	,594**	-,194**	-,916**	,066*	1
	Sig. (2-seitig)	,722	,000	,000	,000	,000	,000	,000	,156	,000	,000	,000	,000	,000	,000	,048	
	N	897	897	897	897	897	897	897	897	897	897	897	897	897	897	897	897

Tabelle 6.3: Auserwählte Modelle für die 15. Bundestagswahl (2002) der Wahlbeteiligung, Erststimme CDU/CSU und Zweistimme CDU/CSU.

Koeffizienten^a

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	t	Sig.	Konfidenzintervall für B (95,0%)		Kollinearitätsstatistik	
		B	Standardfehler	Beta			Untergrenze	Obergrenze	Toleranz	VIF
WB	a	82,361	3,223		25,558	,000	76,019	88,704		
	sex	-,086	,028	-,092	-3,063	,002	-,142	-,031	,943	1,060
	mig	-,173	,033	-,247	-5,190	,000	-,238	-,107	,376	2,662
	une	-,424	,043	-,554	-9,954	,000	-,507	-,340	,275	3,640
	age1	,347	,093	,188	3,736	,000	,164	,530	,334	2,992
	age2	-1,544	,199	-,315	-7,755	,000	-1,936	-1,153	,516	1,940
	age3	,664	,120	,278	5,513	,000	,427	,901	,335	2,988
	edu3	,015	,024	,024	,635	,526	-,032	,063	,588	1,700
	edu4	,106	,026	,172	4,089	,000	,055	,157	,479	2,088
	sec1	-,407	,140	-,133	-2,908	,004	-,682	-,131	,408	2,451
CDU/CSU1	a	103,565	11,404		9,082	,000	81,121	126,010		
	mig	-,421	,095	-,207	-4,412	,000	-,608	-,233	,604	1,655
	une	-,894	,145	-,402	-6,162	,000	-1,180	-,609	,312	3,203
	age4	-,099	,344	-,013	-,287	,775	-,776	,579	,690	1,449
	edu1	-,361	,236	-,080	-1,534	,126	-,825	,102	,488	2,051
	edu3	-,541	,095	-,292	-5,707	,000	-,727	-,354	,509	1,963
	edu4	-,608	,101	-,340	-6,004	,000	-,807	-,409	,414	2,414
	sec4	-,154	,071	-,129	-2,166	,031	-,295	-,014	,376	2,658
CDU/CSU2	a	106,356	7,114		14,950	,000	92,354	120,358		
	sex	,069	,104	,025	,661	,509	-,136	,273	,932	1,073
	mig	-,456	,097	-,220	-4,684	,000	-,647	-,264	,593	1,686
	une	-,824	,148	-,364	-5,582	,000	-1,115	-,534	,308	3,250
	edu1	-,450	,229	-,098	-1,964	,050	-,901	,001	,527	1,897
	edu3	-,637	,095	-,337	-6,683	,000	-,824	-,449	,514	1,944
	edu4	-,661	,099	-,363	-6,650	,000	-,856	-,465	,440	2,274
	sec3	-,377	,113	-,123	-3,337	,001	-,599	-,155	,957	1,045
	sec4	-,129	,072	-,106	-1,787	,075	-,271	,013	,375	2,669

Tabelle 6.4: Auserwählte Modelle für die 15. Bundestagswahl (2002) der Erst- und Zweistimme SPD.

Koeffizienten^a

Modell		Nicht standardisierte Koeffizienten	Standardfehler	Standardisierte Koeffizienten	t	Sig.	Konfidenzintervall für B (95,0%)		Kollinearitätsstatistik	
		B		Beta			Untergrenze	Obergrenze	Toleranz	VIF
SPD1	a	-388,241	82,788		-4,690	,000	-551,189	-225,292		
	sex	-,126	,108	-,056	-1,160	,247	-,339	,088	,921	1,085
	mig	,404	,128	,241	3,155	,002	,152	,656	,367	2,722
	une	,654	,167	,357	3,924	,000	,326	,982	,259	3,857
	age1	4,848	,980	1,100	4,949	,000	2,920	6,777	,044	22,974
	age3	4,950	1,124	,866	4,405	,000	2,738	7,161	,056	17,971
	age4	3,172	,904	,491	3,510	,001	1,393	4,950	,110	9,103
	age5	4,862	,796	1,046	6,107	,000	3,295	6,429	,073	13,637
	edu1	,123	,265	,033	,465	,642	-,398	,644	,424	2,359
	edu3	,380	,099	,249	3,850	,000	,186	,574	,515	1,943
	edu4	,485	,118	,330	4,118	,000	,253	,717	,336	2,976
	sec3	,402	,128	,163	3,147	,002	,151	,654	,802	1,247
SPD2	a	-373,645	71,540		-5,223	,000	-514,453	-232,837		
	sex	-,038	,100	-,019	-,382	,703	-,234	,158	,925	1,081
	mig	,186	,116	,122	1,601	,111	-,043	,414	,379	2,642
	une	,871	,153	,523	5,701	,000	,570	1,171	,261	3,825
	age1	4,791	,860	1,195	5,573	,000	3,099	6,483	,048	20,912
	age3	4,711	,966	,906	4,877	,000	2,809	6,612	,064	15,697
	age4	3,127	,764	,532	4,094	,000	1,624	4,631	,130	7,691
	age5	4,521	,688	1,069	6,570	,000	3,167	5,875	,083	12,042
	edu1	,065	,243	,019	,267	,789	-,414	,544	,425	2,354
	edu3	,413	,091	,297	4,558	,000	,234	,591	,517	1,932
	edu4	,442	,108	,330	4,093	,000	,230	,655	,338	2,962

Tabelle 6.5: Auserwählte Modelle für die 15. Bundestagswahl (2002) der Erst- und Zweistimme FDP und Erststimme der Linken.

Koeffizienten 2002										
Modell	B	Standardfehler	Beta	T	Sig.	95,0% Konfidenzintervalle für B		Toleranz	VIF	
						Untergrenze	Obergrenze			
FDP1	a	9,56	3,41	2,803	0,005	2,848	16,271			
	une	-0,09	0,037	-0,272	0,015	-0,163	-0,018	0,229	4,371	
	age2	-0,269	0,146	-0,126	0,065	-0,556	0,017	0,609	1,642	
	age3	-0,164	0,088	-0,158	0,064	-0,338	0,009	0,393	2,547	
	age4	-0,141	0,082	-0,12	0,086	-0,303	0,02	0,579	1,727	
	edu1	-0,036	0,052	-0,053	0,491	-0,137	0,066	0,485	2,062	
	edu3	0,071	0,021	0,255	0,001	0,029	0,112	0,488	2,051	
	edu4	0,099	0,023	0,37	0	0,054	0,144	0,385	2,6	
	sec1	0,318	0,11	0,239	0,004	0,101	0,535	0,415	2,409	
	sec4	0,033	0,017	0,183	0,052	0	0,066	0,321	3,115	
FDP2	a	23,447	3,67	6,388	0	16,223	30,671			
	mig	0,059	0,021	0,169	0,006	0,017	0,1	0,477	2,096	
	une	-0,079	0,032	-0,208	0,015	-0,143	-0,016	0,244	4,093	
	age1	0,308	0,072	0,336	0	0,167	0,449	0,288	3,467	
	age2	-0,989	0,128	-0,405	0	-1,241	-0,738	0,643	1,555	
	age4	-0,145	0,071	-0,108	0,043	-0,286	-0,004	0,62	1,612	
	edu1	-0,195	0,043	-0,252	0	-0,279	-0,111	0,577	1,732	
	edu2	-0,198	0,021	-0,854	0	-0,24	-0,156	0,204	4,899	
	edu3	-0,083	0,028	-0,261	0,003	-0,137	-0,029	0,234	4,27	
	sec1	0,356	0,103	0,233	0,001	0,152	0,559	0,383	2,613	
LIN1	sec4	0,022	0,015	0,109	0,125	-0,006	0,051	0,35	2,861	
	a	333,737	33,996	9,817	0	266,824	400,65			
	sex	0,028	0,047	0,016	0,549	-0,065	0,121	0,925	1,081	
	mig	-0,16	0,055	-0,117	0,004	-0,269	-0,051	0,379	2,642	
	une	0,37	0,073	0,248	0	0,227	0,513	0,261	3,825	
	age1	-4,116	0,409	-1,147	0	-4,92	-3,312	0,048	20,912	
	age3	-4,696	0,459	-1,009	0	-5,6	-3,793	0,064	15,697	
	age4	-2,833	0,363	-0,539	0	-3,548	-2,118	0,13	7,691	
	age5	-4,162	0,327	-1,099	0	-4,806	-3,518	0,083	12,042	
	edu1	0,448	0,116	0,148	0	0,22	0,675	0,425	2,354	
	edu3	0,082	0,043	0,066	0,058	-0,003	0,167	0,517	1,932	
	edu4	0,134	0,051	0,112	0,01	0,033	0,235	0,338	2,962	

Tabelle 6.6: Auserwählte Modelle für die 15. Bundestagswahl (2002) der Zweitstimme der Linken, Erst- und Zweistimme der Grünen.

Koeffizienten 2002										
Modell		B	Standardfehler	Beta	T	Sig.	95,0% Konfidenzintervalle für B		Toleranz	VIF
							Untergrenze	Obergrenze		
LIN2	a	295,047	25,093		11,758	0	245,658	344,436		
	sex	0,038	0,035	0,025	1,096	0,274	-0,03	0,107	0,925	1,081
	mig	-0,139	0,041	-0,12	-3,421	0,001	-0,219	-0,059	0,379	2,642
	une	0,306	0,054	0,241	5,71	0	0,2	0,411	0,261	3,825
	age1	-3,721	0,302	-1,22	-12,341	0	-4,315	-3,128	0,048	20,912
	age3	-4,152	0,339	-1,05	-12,255	0	-4,819	-3,485	0,064	15,697
	age4	-2,529	0,268	-0,566	-9,439	0	-3,057	-2,002	0,13	7,691
	age5	-3,588	0,241	-1,115	-14,867	0	-4,063	-3,113	0,083	12,042
	edu1	0,372	0,085	0,145	4,364	0	0,204	0,54	0,425	2,354
	edu3	0,092	0,032	0,087	2,908	0,004	0,03	0,155	0,517	1,932
GRÜ1	a	295,047	25,093		11,758	0	245,658	344,436		
	sex	0,038	0,035	0,025	1,096	0,274	-0,03	0,107	0,925	1,081
	mig	-0,139	0,041	-0,12	-3,421	0,001	-0,219	-0,059	0,379	2,642
	une	0,306	0,054	0,241	5,71	0	0,2	0,411	0,261	3,825
	age1	-3,721	0,302	-1,22	-12,341	0	-4,315	-3,128	0,048	20,912
	age3	-4,152	0,339	-1,05	-12,255	0	-4,819	-3,485	0,064	15,697
	age4	-2,529	0,268	-0,566	-9,439	0	-3,057	-2,002	0,13	7,691
	age5	-3,588	0,241	-1,115	-14,867	0	-4,063	-3,113	0,083	12,042
	edu1	0,372	0,085	0,145	4,364	0	0,204	0,54	0,425	2,354
	edu3	0,092	0,032	0,087	2,908	0,004	0,03	0,155	0,517	1,932
GRÜ2	a	-112,31	18,632		-6,028	0	-148,977	-75,636		
	mig	0,179	0,033	0,239	5,406	0	0,114	0,245	0,383	2,61
	age1	0,891	0,22	0,451	4,044	0	0,457	1,325	0,06	16,609
	age3	2,368	0,247	0,923	9,571	0	1,881	2,855	0,08	12,446
	age4	1,249	0,202	0,431	6,201	0	0,853	1,646	0,155	6,466
	age5	0,876	0,181	0,42	4,853	0	0,521	1,231	0,1	10,014
	edu1	-0,024	0,066	-0,015	-0,37	0,712	-0,154	0,106	0,477	2,097
	edu3	-0,008	0,024	-0,012	-0,337	0,736	-0,055	0,039	0,614	1,629
	edu4	0,065	0,032	0,099	2,014	0,045	0,001	0,129	0,312	3,205
	sec4	0,046	0,02	0,105	2,33	0,02	0,007	0,086	0,369	2,708

Tabelle 6.7: Auserwählte Modelle für die 16. Bundestagswahl (2005) der Wahlbeteiligung und Erststimme der CDU/CSU

Koeffizienten^a

Modell		Nicht standardisierte Koeffizienten	Standardfehler	Standardisierte Koeffizienten	t	Sig.	Konfidenzintervall für B (95,0%)		Kollinearitätsstatistik	
		B		Beta			Untergrenze	Obergrenze	Toleranz	VIF
WB	a	102,980	16,037		6,421	,000	71,414	134,546		
	sex	-,476	,244	-,106	-1,953	,052	-,955	,004	,490	2,041
	mig	-,220	,037	-,397	-5,877	,000	-,294	-,146	,317	3,157
	une	-,204	,046	-,358	-4,471	,000	-,293	-,114	,225	4,439
	age2	-1,559	,204	-,476	-7,653	,000	-1,959	-1,158	,374	2,675
	age3	,229	,139	,130	1,652	,100	-,044	,503	,232	4,310
	age4	-,030	,124	-,013	-,240	,810	-,273	,214	,534	1,871
	age5	-,354	,101	-,255	-3,522	,000	-,552	-,156	,276	3,626
	edu2	,197	,062	,567	3,156	,002	,074	,320	,045	22,313
	edu3	,228	,063	,531	3,596	,000	,103	,352	,066	15,087
	edu4	,321	,064	,713	5,054	,000	,196	,447	,073	13,745
	sec1	-,155	,140	-,069	-1,110	,268	-,431	,120	,375	2,668
CDU/CSU1	a	335,357	49,975		6,710	,000	236,992	433,722		
	sex	-3,373	,817	-,210	-4,127	,000	-4,981	-1,764	,491	2,038
	une	-,911	,156	-,448	-5,839	,000	-1,218	-,604	,216	4,620
	age2	-,019	,640	-,002	-,030	,976	-1,279	1,240	,427	2,343
	age3	-1,045	,414	-,166	-2,525	,012	-1,859	-,230	,295	3,395
	age4	-1,333	,419	-,157	-3,181	,002	-2,158	-,508	,524	1,907
	age5	-1,236	,350	-,249	-3,531	,000	-1,925	-,547	,257	3,895
	edu1	-,133	,234	-,028	-,568	,571	-,593	,327	,517	1,936
	edu3	-,265	,081	-,173	-3,256	,001	-,425	-,105	,453	2,208
	edu4	-,329	,107	-,204	-3,068	,002	-,541	-,118	,288	3,475
	sec1	1,172	,455	,146	2,578	,010	,277	2,067	,400	2,501
	sec4	-,205	,075	-,184	-2,725	,007	-,353	-,057	,280	3,572

Tabelle 6.8: Auserwählte Modelle für die 16. Bundestagswahl (2005) der Zweitstimme der CDU/CSU und Erststimme der SPD.

Koeffizienten^a

Modell		Nicht standardisierte Koeffizienten	Standardfehler	Standardisierte Koeffizienten	t	Sig.	Konfidenzintervall für B (95,0%)		Kollinearitätsstatistik	
		B		Beta			Untergrenze	Obergrenze	Toleranz	VIF
CDU/CSU2	a	273,526	45,970		5,950	,000	183,044	364,009		
	sex	-2,698	,739	-,189	-3,653	,000	-4,152	-1,245	,477	2,094
	mig	-,212	,114	-,120	-1,866	,063	-,435	,012	,309	3,234
	une	-,733	,140	-,406	-5,249	,000	-1,007	-,458	,215	4,651
	age2	,130	,597	,013	,218	,827	-1,045	1,306	,390	2,565
	age3	-,484	,455	-,087	-1,064	,288	-1,380	,411	,194	5,166
	age4	-1,023	,388	-,136	-2,637	,009	-1,787	-,259	,486	2,058
	age5	-1,015	,316	-,230	-3,213	,001	-1,638	-,393	,250	3,994
	edu1	-,157	,210	-,038	-,748	,455	-,571	,257	,507	1,972
	edu3	-,299	,073	-,219	-4,105	,000	-,442	-,155	,451	2,219
	edu4	-,377	,096	-,263	-3,941	,000	-,566	-,189	,288	3,476
	sec1	,941	,425	,131	2,211	,028	,103	1,778	,363	2,754
	sec4	-,184	,067	-,186	-2,735	,007	-,317	-,052	,279	3,589
SPD1	a	-465,460	91,930		-5,063	,000	-646,402	-284,518		
	sex	1,973	,979	,137	2,015	,045	,045	3,901	,482	2,075
	mig	,452	,151	,253	3,002	,003	,156	,749	,312	3,207
	age1	5,063	,708	1,182	7,151	,000	3,669	6,456	,081	12,351
	age3	3,535	,953	,624	3,710	,000	1,659	5,411	,078	12,768
	age4	2,960	,785	,386	3,771	,000	1,415	4,505	,211	4,741
	age5	4,990	,704	1,114	7,092	,000	3,605	6,375	,090	11,154
	edu1	,259	,269	,061	,961	,337	-,271	,788	,550	1,819
	edu3	,195	,094	,141	2,075	,039	,010	,380	,479	2,087
	edu4	,400	,127	,275	3,141	,002	,149	,651	,289	3,464
	sec1	-,130	,540	-,018	-,241	,810	-1,193	,933	,400	2,500
	sec4	,387	,082	,385	4,737	,000	,226	,548	,336	2,980

Tabelle 6.9: Auserwählte Modelle für die 16. Bundestagswahl (2005) der Zweitstimme der SPD und Erst- und Zweitstimme der FDP.

Koeffizienten^a

Modell		Nicht standardisierte Koeffizienten	Standardfehler	Standardisierte Koeffizienten	t	Sig.	Konfidenzintervall für B (95,0%)		Kollinearitätsstatistik	
		B		Beta			Untergrenze	Obergrenze	Toleranz	VIF
SPD2	a	-500,621	71,736		-6,979	,000	-641,816	-359,427		
	sex	2,164	,804	,185	2,693	,007	,582	3,746	,492	2,032
	une	,777	,153	,525	5,066	,000	,475	1,078	,217	4,600
	age1	5,473	,623	1,582	8,785	,000	4,247	6,699	,072	13,915
	age3	4,656	,704	1,017	6,617	,000	3,271	6,040	,099	10,127
	age4	3,423	,619	,553	5,530	,000	2,204	4,641	,233	4,288
	age5	4,806	,551	1,328	8,726	,000	3,722	5,890	,101	9,942
	edu1	-,012	,230	-,003	-,051	,960	-,464	,441	,518	1,930
	edu3	,116	,080	,104	1,448	,149	-,042	,274	,452	2,212
	edu4	,269	,106	,229	2,549	,011	,061	,477	,288	3,468
	sec1	-1,418	,448	-,242	-3,168	,002	-2,300	-,537	,400	2,501
	sec4	,204	,074	,251	2,748	,006	,058	,350	,280	3,572
FDP1	a	6,844	10,537		,649	,517	-13,897	27,584		
	sex	-,406	,168	-,186	-2,419	,016	-,737	-,076	,489	2,047
	mig	,025	,026	,091	,947	,345	-,027	,076	,311	3,218
	une	-,174	,029	-,628	-5,961	,000	-,232	-,117	,259	3,861
	age2	,636	,137	,398	4,632	,000	,366	,907	,390	2,565
	age3	,055	,101	,064	,541	,589	-,145	,254	,207	4,842
	age4	,233	,089	,201	2,632	,009	,059	,407	,495	2,022
	age5	,175	,073	,259	2,414	,016	,032	,318	,251	3,988
	edu1	-,027	,048	-,042	-,560	,576	-,122	,068	,509	1,964
	edu3	,033	,017	,155	1,956	,051	,000	,065	,456	2,193
	edu4	-,019	,020	-,087	-,959	,338	-,058	,020	,353	2,834
	sec1	,333	,096	,303	3,455	,001	,143	,523	,374	2,672
FDP2	a	38,922	9,253		4,206	,000	20,710	57,134		
	sex	-,615	,191	-,180	-3,220	,001	-,990	-,239	,602	1,660
	mig	,082	,026	,194	3,139	,002	,031	,133	,494	2,023
	une	-,327	,035	-,756	-9,324	,000	-,396	-,258	,287	3,489
	age2	-,222	,144	-,089	-1,535	,126	-,506	,063	,562	1,781
	edu1	-,039	,056	-,038	-,686	,493	-,149	,072	,600	1,667
	edu3	,082	,021	,251	3,954	,000	,041	,123	,467	2,141
	edu4	,069	,025	,202	2,755	,006	,020	,119	,350	2,858
	sec1	,474	,119	,277	3,987	,000	,240	,708	,391	2,555
	sec4	,017	,018	,073	,944	,346	-,019	,054	,313	3,200

Tabelle 6.10: Auserwählte Modelle für die 16. Bundestagswahl (2005) der Erst- und Zweitstimme der Linken.

Koeffizienten										
Modell		Nicht standardisierte Koeffizienten	Standardfehler	Standardisierte Koeffizienten	t	Sig.	Konfidenzintervall für B (95,0%)		Kollinearitätsstatistik	
		B		Beta			Untergrenze	Obergrenze	Toleranz	VIF
LIN1	a	-267,894	26,828		-9,986	0	-320,698	-215,09		
	sex	2,012	0,428	0,148	4,705	0	1,17	2,854	0,489	2,047
	mig	-0,233	0,066	-0,138	-3,506	0,001	-0,363	-0,102	0,311	3,218
	une	0,342	0,074	0,198	4,588	0	0,195	0,488	0,259	3,861
	age2	4,892	0,35	0,493	13,99	0	4,204	5,58	0,39	2,565
	age3	0,447	0,258	0,084	1,734	0,084	-0,06	0,955	0,207	4,842
	age4	2,372	0,225	0,329	10,524	0	1,928	2,815	0,495	2,022
	age5	1,457	0,185	0,346	7,876	0	1,093	1,821	0,251	3,988
	edu1	0,156	0,123	0,039	1,265	0,207	-0,086	0,397	0,509	1,964
	edu3	0,162	0,042	0,125	3,825	0	0,079	0,245	0,456	2,193
	edu4	0,056	0,051	0,041	1,109	0,268	-0,044	0,156	0,353	2,834
	sec1	-0,291	0,245	-0,043	-1,186	0,236	-0,774	0,192	0,374	2,672
LIN2	a	-108,11	14,383		-7,516	0	-136,42	-79,799		
	sex	1,925	0,342	0,151	5,632	0	1,252	2,598	0,505	1,98
	mig	-0,22	0,052	-0,139	-4,273	0	-0,321	-0,119	0,34	2,937
	une	0,359	0,063	0,221	5,67	0	0,234	0,483	0,237	4,216
	age1	-1,479	0,151	-0,391	-9,827	0	-1,776	-1,183	0,229	4,375
	age2	2,864	0,262	0,307	10,912	0	2,347	3,38	0,457	2,188
	age3	-0,815	0,188	-0,163	-4,336	0	-1,184	-0,445	0,257	3,889
	age4	0,73	0,166	0,108	4,41	0	0,404	1,056	0,605	1,652
	edu1	0,218	0,1	0,058	2,186	0,03	0,022	0,414	0,51	1,96
	edu3	0,152	0,034	0,124	4,46	0	0,085	0,218	0,468	2,139
	edu4	0,089	0,046	0,069	1,953	0,052	-0,001	0,178	0,288	3,466
	sec4	-0,038	0,032	-0,043	-1,207	0,228	-0,1	0,024	0,287	3,481

Tabelle 6.11: Auserwählte Modelle für die 16. Bundestagswahl (2005) der Erst- und Zweitstimme der Grünen.

Koeffizienten										
Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	t	Sig.	Konfidenzintervall für B (95,0%)		Kollinearitätsstatistik	
		B	Standardfehler	Beta			Untergrenze	Obergrenze	Toleranz	VIF
GRÜ1	a	142,435	19,203		7,417	0	104,64	180,231		
	sex	-0,624	0,311	-0,117	-2,006	0,046	-1,235	-0,012	0,503	1,988
	une	0,081	0,055	0,12	1,475	0,141	-0,027	0,189	0,26	3,847
	age1	-1,273	0,154	-0,806	-8,277	0	-1,576	-0,971	0,18	5,549
	age2	-1,371	0,276	-0,352	-4,965	0	-1,914	-0,827	0,34	2,939
	age4	-0,859	0,168	-0,304	-5,104	0	-1,19	-0,528	0,483	2,071
	age5	-1,606	0,134	-0,973	-11,969	0	-1,87	-1,342	0,259	3,859
	edu1	0,055	0,09	0,035	0,612	0,541	-0,121	0,231	0,522	1,916
	edu3	-0,039	0,031	-0,076	-1,239	0,216	-0,1	0,023	0,458	2,183
	edu4	-0,063	0,037	-0,117	-1,684	0,093	-0,136	0,011	0,354	2,828
GRÜ2	sec1	-0,132	0,173	-0,049	-0,762	0,447	-0,472	0,208	0,411	2,435
	a	15,977	9,301		1,718	0,087	-2,331	34,285		
	sex	-1,002	0,219	-0,182	-4,581	0	-1,433	-0,571	0,516	1,94
	mig	0,127	0,034	0,185	3,733	0	0,06	0,193	0,328	3,049
	age1	0,215	0,09	0,132	2,4	0,017	0,039	0,392	0,269	3,716
	age2	-0,155	0,172	-0,038	-0,898	0,37	-0,494	0,184	0,443	2,256
	age3	1,28	0,113	0,591	11,324	0	1,057	1,502	0,297	3,365
	age4	0,564	0,108	0,193	5,207	0	0,351	0,777	0,592	1,69
	edu1	-0,023	0,061	-0,014	-0,381	0,703	-0,144	0,098	0,562	1,78
	edu3	-0,049	0,021	-0,094	-2,38	0,018	-0,09	-0,009	0,523	1,91
	edu4	0,055	0,029	0,099	1,861	0,064	-0,003	0,113	0,288	3,469
	sec3	0,017	0,029	0,018	0,595	0,553	-0,04	0,074	0,841	1,189
sec4	0,021	0,019	0,055	1,104	0,271	-0,016	0,058	0,332	3,015	

Tabelle 6.12: Auserwählte Modelle für die 17. Bundestagswahl (2009) der Wahlbeteiligung, Erst- und Zweitstimme CDU/CSU.

Modell		B	Standardfehler	Beta	t	Sig.	Konfidenzintervall für B (95,0%)		Toleranz	VIF
							Untergrenze	Obergrenze		
WB	a	164,476	19,572		8,403	0	125,953	202,999		
	sex	-1,668	0,301	-0,254	-5,538	0	-2,261	-1,076	0,699	1,43
	une	-0,468	0,087	-0,412	-5,403	0	-0,639	-0,298	0,253	3,956
	age1	0,764	0,181	0,414	4,221	0	0,408	1,121	0,153	6,541
	age2	-1,188	0,291	-0,208	-4,082	0	-1,761	-0,615	0,566	1,767
	age3	0,316	0,166	0,144	1,91	0,057	-0,01	0,642	0,259	3,861
	age5	-0,245	0,178	-0,123	-1,372	0,171	-0,595	0,106	0,183	5,456
	edu1	0,059	0,129	0,023	0,456	0,649	-0,196	0,314	0,556	1,798
	edu2	-0,151	0,041	-0,255	-3,651	0	-0,232	-0,069	0,301	3,323
	edu3	-0,127	0,043	-0,174	-2,987	0,003	-0,211	-0,043	0,432	2,316
CDU/CSU1	a	94,165	27,385		3,439	0,001	40,267	148,063		
	sex	-0,876	0,556	-0,069	-1,575	0,116	-1,971	0,219	0,755	1,325
	mig	-0,296	0,08	-0,185	-3,684	0	-0,453	-0,138	0,572	1,748
	une	-0,937	0,161	-0,426	-5,814	0	-1,254	-0,62	0,269	3,716
	age4	-0,329	0,287	-0,05	-1,147	0,252	-0,894	0,236	0,749	1,335
	age5	-0,281	0,199	-0,073	-1,408	0,16	-0,673	0,112	0,539	1,854
	edu1	-0,459	0,245	-0,094	-1,875	0,062	-0,94	0,023	0,573	1,744
	edu2	0,267	0,069	0,234	3,873	0	0,131	0,402	0,397	2,52
	edu3	0,375	0,074	0,265	5,042	0	0,228	0,521	0,524	1,908
CDU/CSU2	a	59,603	11,899		5,009	0	36,184	83,022		
	mig	-0,399	0,065	-0,309	-6,187	0	-0,526	-0,272	0,576	1,737
	une	-0,849	0,118	-0,476	-7,172	0	-1,082	-0,616	0,325	3,079
	age1	-0,448	0,219	-0,155	-2,045	0,042	-0,879	-0,017	0,251	3,984
	age4	-0,479	0,226	-0,09	-2,117	0,035	-0,924	-0,034	0,786	1,273
	age5	-0,465	0,172	-0,149	-2,706	0,007	-0,802	-0,127	0,474	2,11
	edu2	0,243	0,062	0,263	3,945	0	0,122	0,365	0,322	3,103
	edu3	0,402	0,065	0,351	6,203	0	0,275	0,53	0,448	2,23

Tabelle 6.13: Auserwählte Modelle für die 17. Bundestagswahl (2009) der Erst- und Zweitstimme SPD und Erststimme FDP.

Modell		B	Standardfehler	Beta	T	Sig.	Konfidenzintervall für B (95,0%)		Toleranz	VIF
							Untergrenze	Obergrenze		
SPD1	a	403,232	41,48		9,721	0	321,59	484,875		
	sex	-2,398	0,72	-0,182	-3,331	0,001	-3,814	-0,981	0,685	1,46
	mig	0,72	0,128	0,434	5,639	0	0,468	0,971	0,343	2,913
	une	0,773	0,208	0,339	3,719	0	0,364	1,182	0,245	4,075
	age2	-3,134	0,812	-0,273	-3,859	0	-4,733	-1,536	0,406	2,46
	age3	-4,081	0,448	-0,924	-9,104	0	-4,963	-3,198	0,198	5,056
	age4	-3,11	0,449	-0,459	-6,928	0	-3,994	-2,227	0,464	2,153
	age5	-2,406	0,333	-0,602	-7,232	0	-3,061	-1,751	0,294	3,4
	edu1	0,832	0,312	0,165	2,666	0,008	0,218	1,446	0,534	1,871
	edu2	-0,478	0,097	-0,403	-4,905	0	-0,67	-0,286	0,301	3,32
SPD2	edu3	-0,313	0,103	-0,213	-3,049	0,003	-0,515	-0,111	0,416	2,401
	a	301,397	30,326		9,938	0	241,708	361,087		
	sex	-1,38	0,526	-0,14	-2,623	0,009	-2,416	-0,345	0,685	1,46
	mig	0,466	0,093	0,377	4,999	0	0,283	0,65	0,343	2,913
	une	0,762	0,152	0,447	5,012	0	0,463	1,061	0,245	4,075
	age2	-2,559	0,594	-0,299	-4,31	0	-3,728	-1,391	0,406	2,46
	age3	-3,349	0,328	-1,016	-10,219	0	-3,994	-2,704	0,198	5,056
	age4	-2,662	0,328	-0,526	-8,111	0	-3,308	-2,016	0,464	2,153
	age5	-1,843	0,243	-0,618	-7,576	0	-2,322	-1,364	0,294	3,4
	edu1	0,618	0,228	0,164	2,709	0,007	0,169	1,067	0,534	1,871
FDP1	edu2	-0,34	0,071	-0,384	-4,773	0	-0,48	-0,2	0,301	3,32
	edu3	-0,238	0,075	-0,217	-3,164	0,002	-0,385	-0,09	0,416	2,401
	a	15,192	10,786		1,408	0,16	-6,036	36,42		
	sex	-0,389	0,187	-0,106	-2,08	0,038	-0,757	-0,021	0,915	1,093
	une	-0,277	0,061	-0,437	-4,506	0	-0,397	-0,156	0,254	3,936
	age1	-0,106	0,096	-0,103	-1,107	0,269	-0,295	0,083	0,275	3,637
	edu2	0,215	0,072	0,653	2,974	0,003	0,073	0,357	0,05	20,183
	edu3	0,183	0,067	0,45	2,735	0,007	0,051	0,315	0,088	11,359
	edu4	0,18	0,074	0,533	2,437	0,015	0,035	0,325	0,05	20,036

Tabelle 6.14: Auserwählte Modelle für die 17. Bundestagswahl (2009) der Zweitstimme FDP und Erststimme der Linken.

Modell		Nicht standardisierte Koeffizienten	Standardfehler	Standardisierte Koeffizienten	t	Sig.	Konfidenzintervall für B (95,0%)		Kollinearitätsstatistik	
		B		Beta			Untergrenze	Obergrenze	Toleranz	VIF
FDP2	a	71,584	14,655		4,885	0	42,741	100,428		
	sex	-0,799	0,247	-0,158	-3,24	0,001	-1,284	-0,314	0,698	1,432
	mig	0,129	0,043	0,203	2,991	0,003	0,044	0,215	0,357	2,801
	une	-0,389	0,073	-0,444	-5,298	0	-0,534	-0,244	0,235	4,247
	age2	-0,775	0,291	-0,176	-2,661	0,008	-1,348	-0,202	0,379	2,642
	age3	-0,353	0,155	-0,208	-2,28	0,023	-0,657	-0,048	0,199	5,028
	age4	-0,358	0,163	-0,137	-2,198	0,029	-0,679	-0,037	0,422	2,367
	age5	-0,29	0,121	-0,189	-2,392	0,017	-0,528	-0,051	0,265	3,769
	edu2	0,165	0,085	0,363	1,954	0,052	-0,001	0,332	0,048	20,878
	edu3	0,149	0,081	0,264	1,833	0,068	-0,011	0,309	0,08	12,569
LIN1	edu4	0,214	0,087	0,458	2,452	0,015	0,042	0,386	0,047	21,118
	a	-76,028	27,578		-2,757	0,006	-130,307	-21,749		
	sex	3,17	0,421	0,227	7,526	0	2,341	3,999	0,697	1,435
	mig	-0,576	0,068	-0,329	-8,45	0	-0,71	-0,442	0,419	2,386
	une	0,505	0,12	0,209	4,198	0	0,268	0,742	0,256	3,914
	age1	-2,515	0,259	-0,641	-9,712	0	-3,024	-2,005	0,146	6,857
	age3	-0,363	0,288	-0,078	-1,261	0,208	-0,93	0,204	0,167	5,985
	age5	-0,505	0,241	-0,119	-2,101	0,037	-0,979	-0,032	0,196	5,098
	edu1	-0,06	0,182	-0,011	-0,327	0,744	-0,417	0,298	0,549	1,821
	edu2	-0,101	0,058	-0,081	-1,761	0,079	-0,215	0,012	0,3	3,329
	edu3	-0,134	0,061	-0,087	-2,219	0,027	-0,254	-0,015	0,417	2,399

Tabelle 6.15: Auserwählte Modelle für die 17. Bundestagswahl (2009) der Zweitstimme der Linken, Erst und Zweitstimmen der Grünen.

Modell		B	Standardfehler	Beta	t	Sig.	Konfidenzintervall für B (95,0%)		Kollinearitätsstatistik Toleranz	VIF
							Untergrenze	Obergrenze		
LIN2	a	-56,273	22,419		-2,51	0,013	-100,398	-12,147		
	sex	2,776	0,342	0,215	8,106	0	2,102	3,45	0,697	1,435
	mig	-0,402	0,055	-0,248	-7,248	0	-0,511	-0,293	0,419	2,386
	une	0,511	0,098	0,229	5,227	0	0,319	0,704	0,256	3,914
	age1	-2,523	0,21	-0,696	-11,985	0	-2,937	-2,109	0,146	6,857
	age3	-0,635	0,234	-0,147	-2,714	0,007	-1,096	-0,175	0,167	5,985
	age5	-0,497	0,196	-0,127	-2,541	0,012	-0,882	-0,112	0,196	5,098
	edu1	-0,015	0,148	-0,003	-0,102	0,919	-0,306	0,276	0,549	1,821
	edu2	-0,095	0,047	-0,082	-2,035	0,043	-0,188	-0,003	0,3	3,329
	edu3	-0,098	0,049	-0,068	-1,985	0,048	-0,195	-0,001	0,417	2,399
GRÜ1	a	-2,382	9,455		-0,252	0,801	-20,992	16,228		
	mig	0,195	0,059	0,222	3,297	0,001	0,079	0,311	0,427	2,343
	age1	0,098	0,208	0,05	0,471	0,638	-0,311	0,507	0,173	5,77
	age3	0,921	0,23	0,393	3,996	0	0,467	1,374	0,2	5,007
	age5	-0,01	0,184	-0,005	-0,054	0,957	-0,372	0,352	0,257	3,888
	edu1	-0,152	0,14	-0,057	-1,085	0,279	-0,429	0,124	0,704	1,421
	edu2	0,105	0,044	0,167	2,38	0,018	0,018	0,192	0,393	2,542
	edu3	-0,089	0,05	-0,115	-1,773	0,077	-0,188	0,01	0,462	2,163
GRÜ2	a	77,761	14,897		5,22	0	48,44	107,081		
	sex	-0,758	0,262	-0,12	-2,893	0,004	-1,274	-0,242	0,685	1,46
	mig	0,187	0,046	0,237	4,076	0	0,097	0,277	0,352	2,84
	age2	-0,816	0,284	-0,149	-2,872	0,004	-1,375	-0,257	0,441	2,27
	age3	0,664	0,163	0,315	4,071	0	0,343	0,985	0,198	5,054
	age4	-0,427	0,16	-0,132	-2,661	0,008	-0,742	-0,111	0,483	2,072
	age5	-0,454	0,109	-0,238	-4,15	0	-0,669	-0,239	0,361	2,771
	edu1	-0,119	0,101	-0,049	-1,184	0,238	-0,317	0,079	0,681	1,468
	edu2	0,018	0,032	0,033	0,583	0,56	-0,044	0,08	0,381	2,623
	edu3	-0,109	0,037	-0,156	-2,978	0,003	-0,181	-0,037	0,435	2,301

Tabelle 6.16: Auserwählte Modelle für die 18. Bundestagswahl (2013) der Wahlbeteiligung, Erst- und Zweitstimmen der CDU/CSU.

Koeffizienten ^a										
Modell		B	Standardfehler	Beta	T	Sig.	95,0% Konfidenzintervalle für B		Toleranz	VIF
							Untergrenze	Obergrenze		
WB	a	30,156	17,562		1,717	,087	-4,409	64,721		
	sex	,566	,337	,096	1,680	,094	-,097	1,229	,790	1,266
	une	,343	,095	,258	3,595	,000	,155	,531	,498	2,010
	age1	,474	,168	,218	2,819	,005	,143	,805	,429	2,332
	age5	,230	,100	,151	2,302	,022	,033	,427	,593	1,685
	edu3	-,305	,046	-,385	-6,642	,000	-,395	-,214	,763	1,310
	sec1	-,261	,224	-,073	-1,167	,244	-,702	,179	,652	1,533
	sec3	,140	,054	,139	2,615	,009	,035	,246	,914	1,094
SDU/CSU1	a	-170,744	59,820		-2,854	,005	-288,483	-53,005		
	sex	2,353	,852	,176	2,760	,006	,675	4,030	,718	1,393
	une	-,500	,292	-,167	-1,710	,088	-1,075	,075	,308	3,248
	age1	1,112	,613	,227	1,815	,071	-,094	2,318	,188	5,327
	age2	2,092	1,061	,228	1,972	,050	,004	4,181	,219	4,567
	age4	,628	,605	,089	1,038	,300	-,563	1,818	,394	2,535
	age5	1,088	,446	,317	2,438	,015	,210	1,967	,173	5,778
	edu1	-,960	,396	-,301	-2,423	,016	-1,740	-,180	,189	5,282
	edu2	-,589	,353	-,173	-1,670	,096	-1,284	,105	,274	3,654
	edu4	,427	,112	,292	3,792	,000	,205	,648	,493	2,027
	sec4	,308	,096	,344	3,201	,002	,119	,497	,254	3,943
	a	-156,021	50,010		-3,120	,002	-254,453	-57,589		
CDU/CSU2	sex	2,291	,713	,203	3,215	,001	,888	3,693	,718	1,393
	une	-,470	,244	-,185	-1,925	,055	-,951	,011	,308	3,248
	age1	1,060	,512	,255	2,068	,040	,051	2,068	,188	5,327
	age2	1,878	,887	,242	2,116	,035	,131	3,624	,219	4,567
	age4	,541	,506	,091	1,070	,286	-,454	1,536	,394	2,535
	age5	,963	,373	,332	2,580	,010	,228	1,697	,173	5,778
	edu1	-,924	,331	-,343	-2,789	,006	-1,576	-,272	,189	5,282
	edu2	-,692	,295	-,240	-2,345	,020	-1,272	-,111	,274	3,654
	edu4	,332	,094	,269	3,529	,000	,147	,517	,493	2,027
	sec4	,252	,080	,332	3,127	,002	,093	,410	,254	3,943

Tabelle 6.17: Auserwählte Modelle für die 18. Bundestagswahl (2013) der Erst- und Zweitstimmen der SPD.

Koeffizienten ^a										
Modell		95,0% Konfidenzintervalle für B								
		B	Standardfehler	Beta	T	Sig.	Untergrenze	Obergrenze	Toleranz	VIF
SPD1	(Konstante)	-166,076	66,729		-2,489	,013	-297,418	-34,734		
	mig	-,334	,113	-,193	-2,944	,004	-,557	-,111	,396	2,526
	une	2,345	,246	,742	9,548	,000	1,861	2,828	,281	3,554
	age1	3,929	,826	,759	4,759	,000	2,304	5,554	,067	14,991
	age3	2,691	,947	,623	2,842	,005	,827	4,554	,035	28,281
	age4	1,711	,684	,231	2,502	,013	,365	3,057	,199	5,027
	age5	1,691	,687	,467	2,462	,014	,339	3,042	,047	21,213
	edu1	,661	,331	,197	1,998	,047	,010	1,312	,175	5,710
	edu2	-,468	,295	-,130	-1,586	,114	-1,050	,113	,252	3,969
	edu3	-,571	,101	-,303	-5,644	,000	-,770	-,372	,589	1,697
	edu4	-,462	,097	-,300	-4,775	,000	-,652	-,271	,431	2,321
	sec3	,274	,109	,114	2,515	,012	,060	,489	,829	1,207
	sec4	-,080	,078	-,084	-1,022	,308	-,233	,074	,249	4,023
SPD2	(Konstante)	-75,274	60,187		-1,251	,212	-193,740	43,192		
	sex	-,855	,520	-,076	-1,644	,101	-1,879	,168	,728	1,373
	mig	-,192	,082	-,140	-2,333	,020	-,355	-,030	,438	2,282
	une	1,748	,158	,695	11,067	,000	1,437	2,059	,399	2,508
	age1	2,975	,636	,723	4,679	,000	1,724	4,226	,066	15,172
	age3	1,982	,725	,577	2,734	,007	,555	3,409	,035	28,314
	age4	1,350	,524	,229	2,579	,010	,320	2,381	,199	5,029
	age5	1,173	,536	,407	2,186	,030	,117	2,229	,045	22,093
	edu1	,609	,236	,228	2,585	,010	,145	1,073	,202	4,946
	edu3	-,481	,077	-,321	-6,210	,000	-,633	-,328	,589	1,697
	edu4	-,444	,074	-,363	-5,984	,000	-,590	-,298	,428	2,337
	sec3	,208	,084	,109	2,486	,013	,043	,373	,825	1,213
	sec4	-,118	,061	-,157	-1,929	,055	-,238	,002	,237	4,211

Tabelle 6.18: Auserwählte Modelle für die 18. Bundestagswahl (2013) der Erst- und Zweitstimmen der FDP und Erststimme der Linken.

Modell		B	Standardfehler	Beta	T	Sig.	95,0% Konfidenzintervalle für B		Toleranz	VIF
							Untergrenze	Obergrenze		
FDP1	a	1,035	0,883		1,172	0,242	-0,704	2,774		
	age3	-0,043	0,038	-0,1	-1,127	0,261	-0,118	0,032	0,402	2,486
	edu1	-0,042	0,022	-0,124	-1,929	0,055	-0,084	0,001	0,761	1,315
	edu4	0,032	0,011	0,21	2,857	0,005	0,01	0,055	0,58	1,723
	sec4	0,022	0,009	0,235	2,525	0,012	0,005	0,039	0,364	2,749
FDP2	a	-8,029	6,211		-1,293	0,197	-20,254	4,196		
	sex	0,173	0,121	0,075	1,424	0,156	-0,066	0,411	0,844	1,185
	mig	-0,061	0,017	-0,217	-3,555	0	-0,095	-0,027	0,632	1,583
	une	0,126	0,031	0,244	4,023	0	0,064	0,187	0,641	1,561
	edu3	-0,047	0,018	-0,154	-2,627	0,009	-0,083	-0,012	0,68	1,47
	edu4	0,064	0,016	0,255	3,99	0	0,032	0,096	0,574	1,741
	sec3	0,026	0,019	0,067	1,356	0,176	-0,012	0,065	0,957	1,045
	sec4	0,038	0,012	0,247	3,096	0,002	0,014	0,062	0,368	2,719
	a	100,388	33,738		2,976	0,003	33,981	166,794		
	sex	-1,463	0,671	-0,119	-2,182	0,03	-2,784	-0,143	0,7	1,428
LIN1	mig	0,245	0,11	0,162	2,217	0,027	0,027	0,462	0,391	2,558
	une	-0,991	0,239	-0,359	-4,155	0	-1,461	-0,522	0,279	3,58
	age1	-1,266	0,365	-0,28	-3,47	0,001	-1,984	-0,548	0,32	3,123
	age2	1,038	0,614	0,123	1,69	0,092	-0,171	2,246	0,395	2,53
	age3	0,006	0,381	0,001	0,015	0,988	-0,744	0,755	0,205	4,878
	age4	0,074	0,425	0,011	0,175	0,861	-0,762	0,911	0,483	2,071
	edu2	0,355	0,271	0,113	1,313	0,19	-0,177	0,888	0,281	3,555
	edu3	0,625	0,095	0,38	6,609	0	0,439	0,811	0,631	1,584
	edu4	-0,086	0,086	-0,064	-0,999	0,319	-0,256	0,084	0,508	1,969
	sec3	-0,255	0,105	-0,121	-2,435	0,015	-0,462	-0,049	0,839	1,192
	sec4	-0,241	0,075	-0,291	-3,19	0,002	-0,389	-0,092	0,25	4,005

Tabelle 6.19: Auserwählte Modelle für die 18. Bundestagswahl (2013) der Zweitstimme der Linken Erst- und Zweitstimmen der Grünen.

Modell		B	Standardfehler	Beta	T	Sig.	95,0% Konfidenzintervalle für B		Toleranz	VIF
							Untergrenze	Obergrenze		
LIN2	a	82,832	29,353		2,822	0,005	25,057	140,606		
	sex	-1,172	0,584	-0,109	-2,008	0,046	-2,321	-0,023	0,7	1,428
	mig	0,209	0,096	0,159	2,176	0,03	0,02	0,398	0,391	2,558
	une	-0,828	0,208	-0,344	-3,988	0	-1,236	-0,419	0,279	3,58
	age1	-1,042	0,317	-0,265	-3,282	0,001	-1,666	-0,417	0,32	3,123
	age2	0,863	0,534	0,117	1,616	0,107	-0,188	1,914	0,395	2,53
	age3	-0,011	0,331	-0,003	-0,035	0,972	-0,663	0,64	0,205	4,878
	age4	0,098	0,37	0,017	0,266	0,79	-0,629	0,826	0,483	2,071
	edu2	0,344	0,235	0,126	1,463	0,144	-0,119	0,808	0,281	3,555
	edu3	0,556	0,082	0,388	6,754	0	0,394	0,718	0,631	1,584
	edu4	-0,097	0,075	-0,083	-1,3	0,195	-0,245	0,05	0,508	1,969
	sec3	-0,257	0,091	-0,14	-2,817	0,005	-0,437	-0,077	0,839	1,192
GRÜ1	sec4	-0,203	0,066	-0,282	-3,09	0,002	-0,332	-0,074	0,25	4,005
	a	13,301	2,395		5,553	0	8,587	18,014		
	une	-0,227	0,111	-0,17	-2,044	0,042	-0,445	-0,008	0,434	2,303
	age2	-0,482	0,241	-0,118	-1,997	0,047	-0,957	-0,007	0,859	1,164
	edu2	0,383	0,126	0,253	3,043	0,003	0,135	0,631	0,435	2,298
GRÜ2	edu3	-0,17	0,048	-0,215	-3,54	0	-0,265	-0,076	0,818	1,222
	a	12,717	2,488		5,111	0	7,82	17,614		
	mig	-0,042	0,037	-0,067	-1,137	0,256	-0,115	0,031	0,814	1,229
	age2	-0,343	0,214	-0,097	-1,602	0,11	-0,764	0,078	0,762	1,312
	edu2	0,145	0,079	0,111	1,84	0,067	-0,01	0,3	0,776	1,289
	edu3	-0,201	0,039	-0,293	-5,203	0	-0,277	-0,125	0,886	1,129
	sec3	0,066	0,048	0,076	1,394	0,164	-0,027	0,16	0,956	1,046

Erklärung zur Urheberschaft

Hiermit erkläre ich, Huong Viet Pham, dass ich die vorliegende Arbeit allein und unter Verwendung der aufgeführten Quellen und Hilfsmittel angefertigt habe. Die Prüfungsordnung ist mir bekannt. Ich habe in meinem Studienfach bisher keine Bachelorarbeit eingereicht und diese bzw. diese nicht endgültig nicht bestanden.

.....

Huong Viet Pham

Berlin, den 26. Januar 2015